ARTICULATION OF WHISPERED ALVEOLAR CONSONANTS

BY

MEGAN JO OSFAR

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Adviser:

Assistant Professor Ryan K. Shosted

ABSTRACT

This thesis explores the articulation of whispered and normally-spoken alveolar consonants through the use of electropalatography (EPG). While whisper is a well-researched topic, most prior studies have been limited to glottal and acoustic examinations. This study aims to directly measure the articulatory changes speech undergoes when it is whispered. Alveolar consonants /t/, /d/, and /n/ were studied in /iCi/, /aCa/, and /uCu/ environments. Maximum center of gravity, average center of gravity, contacted surface area, and contacted surface area variability were measured. Whispered consonants were found to have a greater duration and COG than normal consonants. Whispered consonants also showed evidence of hyperarticulation: they had lower surface area contact percentages and variability versus normal speech.

ACKNOWLEDGMENTS

Thank you (*in order of appearance*):

Freyja, The Amazing Hiking Cat

Ryan Shosted, Advisor Extraordinaire

Daniel Scarpace, Master of Goats

Charlos, Finest Hound Dog in the (Mid)West

Zach Claybaugh, my patient and amazing partner in crime

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Whisper is a mode of speech production characterized by a lack of regular (modal) vibration at the vocal folds. Typically produced with a combination of relatively low transglottal air flow and minimal glottal impedance, whispered speech is most often aperiodic and low in acoustic energy. Despite these limitations, whispered speech is still largely capable of conveying linguistic information to a listener (Lim, 2010; Mills, 2009; Dannenbring, 1980; Tartter, 1989).

While whisper is only used paralinguistically, its usage can affect the implementation of voicing, an important phonological contrast. Since the phonological feature [+VOICE] is predicated on the vibration of the vocal folds (Chomsky and Halle, 1968, among others), the absence of vocal fold vibration during (most types of) whisper makes the distinction between voiced and voiceless sounds problematic (Mills, 2003).[1] It has been claimed by a variety of researchers that suparlaryngeal adjustment is associated with voicing contrasts in consonants (Ohala, 1986; Hoole, 1998; Higashikawa, et al. 2003; Jovičić and Šarić, 2008; Fuchs et al., 2007; McLeod et al., 2003; Yoshioka, 2008). It is not clear, however, whether the phonological distinction between voiced and voiceless consonants is maintained in whispered speech. I will examine whether suparlaryngeal articulation helps differentiate phonologically voiced and voiceless stops in whispered speech, i.e., when (phonetic) voicing is rendered impossible.

1.1 The communicative function of whispered speech

Whisper is not used for contrastive linguistic purposes (Laver, 1980: 121). However, speakers whisper for a variety of paralinguistic reasons. For example, when a patient suffers from laryngitis or has undergone trauma to the larynx, perhaps from surgical intubation, removal of a cyst, or even removal of one or both vocal folds, whispering may be the only mode of verbal communication possible (without an electronic prosthesis). Whispering is sometimes used to increase comprehensibility, as it has been shown to increase the fluency of speech in adults who

[1] Whisper also necessarily complicates the speaker's ability to communicate prosodic content like stress, intonation, and lexical tone, a relatively better developed area of research on the phonological implications of whisper (Chang & Yao 2007; Liu and Samuel 2004; Nicholson and Teig 2003; Tartter 1994).

stutter (Ingham et al., 2009). Whisper is also used as an aid in the teaching of L2 consonant production: students can be trained to produce unfamiliar phonologically voiced consonants by practicing the whispered version of the phonologically unvoiced counterparts (Tanokuchi et al., 1986).

Jovičić and Šarić (2008) mention two social purposes for whisper. In certain social situations, one may wish to speak in a quiet voice while preserving intelligibility, as in a library. In this case, the speaker chooses whisper to observe a social convention. Because the content of the speech is not meant to be encrypted, intelligibility is preserved despite the drop in amplitude. One may also whisper in order to obfuscate sensitive information. In this case, the speaker may desire some loss of intelligibility or distortion of the speech signal, in consideration of potential eavesdroppers. These are both examples of a "low-energy whisper" (LEW), as defined by Solomon et al. (1989).

Another type of whisper is intended neither to reduce loudness nor to encrypt content: the stage whisper. When an actor whispers on stage with the intent to be heard by the audience, he or she communicates that his or her character's speech is not audible to one or more of the other characters portrayed on stage (an "aside"). The voice still lacks active vocal cord vibration (passive vibration can sometimes result), but is highly intelligible at a distance and requires more pulmonic force and, consequently, air flow. Solomon et al. (1989) call this "high-energy whisper" (HEW).

This paper focuses on the production of intelligible LEW. To my knowledge, no studies offer a comparison of the supralaryngeal articulation of intelligible and encrypted LEW. Investigating other varieties of whisper (i.e, HEW and encrypted LEW) falls beyond the scope of this paper, though such investigations are recommended in order to better understand the full range of what is meant by "whispered speech" (defined narrowly here as intelligible LEW).

Minimal pairs that differ in terms of phonological voice, e.g. *dog* /dag/ and *talk* /tak/ can be produced and perceived in whispered speech, even though the vocal folds do not vibrate (Dannenbring, 1980; Tartter, 1989; Mills, 2003, 2009). Thus, it is necessary to differentiate between the phonological terms "voiced" and "voiceless" (regarding the speaker's intended, phonemic voicing status) and the phonetic terms "normal" and "whispered" (regarding the phonetic state of the vocal folds). I will refer to the voiced/voiceless distinction as "phonological voicing" and the normal/whispered distinction as "speech mode" (following Weismer and Longstreth, 1980, among others). Though this terminology is necessary to frame the goals of this study, it is not altogether adequate. While whisper is a style or mode of speech, it also comprises a range of phonatory settings analogous to the phonatory settings for voicing and voicelessness (Laver, 1980). Furthermore, "normal" speech mode corresponds to at least two phonatory settings, one associated with modal vocal fold vibration (voicing) and one associated with the lack thereof (voicelessness). Based on a review of the evidence (Section 1.3), it seems reasonable to conclude that during whispered speech mode (as operationally defined for present purposes) the vocal folds have only one primary physical setting, which most typically suppresses vibration. Crucially, both phonemic "voiced" and "voiceless" consonants can be produced (and perceived) under this single phonatory regime.[2]

1.2 Perception of phonological voicing in whispered speech

While the basic phonatory mechanism for voicing is abrogated during whispered speech, the categorical distinction between voiced and voiceless consonants is not entirely lost. A number of studies demonstrate that listeners perform at better-than-chance levels when asked to identify the place, manner, and phonemic voicing of whispered consonants, both in forced-choice and open-choice experiments (Dannenbring, 1980; Tartter, 1989; Mills, 2009; Lim, 2010). It has also been observed that some consonant pairs are more easily discriminated than other (Dannenbring, 1980; Tartter, 1989) and that some manners of articulation (fricatives, plosives, etc.) are more easily confused than others (Tartter, 1989; Mills, 2009).

Dannenbring's (1980) perceptual discrimination study of whispered voiced/voiceless consonant pairs uncovered a number of interesting results. Twelve naive university students participated in a forced-choice identification task including pairs of whispered consonants followed by /a/, /i/,

[2]The term "whisper" is used for a variety of purposes in the literature on voice, including as a phonatory counterpart of "voiced" and "voiceless". Laver (1980) presents a helpful overview of the various senses of "whisper", "whisperiness", and "whispery voice".

and /u/ (for example, /pu/ and /bu/). The pairs tested were /b, p/, /d, t/, /g, k/, /z, s/, /v, f/, and /ð, θ/. The subjects were also required to give a confidence rating of 1-14: 1 meant the participant was very sure the consonant was voiceless, 14 meant he or she was sure the consonant was voiced, and 7 meant he or she was unsure of the voicing. The results showed that some pairs (/d, t/, /g, k/) were more easily discriminated than others (/ð, θ/, /z, s/) and that the interaction between consonant pairs and vowels was significant. However, no main effect for vowels was reported.

Tartter's (1989) open set identification study of the perception of whispered consonants, which included a component of acoustic analysis, showed that manner, place of articulation, and voicing information are all transmitted during whispered speech at a level greater than chance. For Tartter's subjects, voicing was the contrast that was hardest for listeners to distinguish (i.e., it had the lowest rate of correct identification).

Examining the acoustic data, Tartter observed that whispered fricatives were much longer when voiceless than voiced. Among the plosives, no significant differences in aspiration levels were observed in the voiced/voiceless pairs. F1 cutback, which has been considered a reliable cue of voicing identification between consonant pairs in whispered speech, was not observed by Tartter in either plosives or fricatives. However, in plosives, a double-spiked burst was observed when the consonant was voiceless. This could serve as a more reliable cue, as burst intensity level often provides information about the presence or absence of voicing in both normal and whispered speech.

Mills (2009) provided further evidence that it is possible to discriminate phonologically voiced from voiceless obstruents in whisper. He presented 288 stimulus pairs (p. 82) to listeners that varied only in the phonological voicing status of the first consonant (i.e., "beer again" vs. "peer again"). The tokens were modified to remove durational cues,[3] a supralaryngeal factor which has been credited with aiding listener discrimination (Mills, 2003). Even with the duration held constant, subjects were able to achieve 75% accuracy on plosives and 56% accuracy on

[3]Voiceless obstruents have a significantly longer closure than voiced obstruents (Tartter, 1989; Mills, 2003).

fricatives, both significantly above chance (p. 84). Mills (2009) conjectures that laryngeal or glottal articulation provides voicing cues in whispered speech.

Lim (2010) found an even greater level of accuracy than Mills (2009) when studying minimal pairs that differ by word-initial consonant. In Lim's study, six English words were presented visually on a computer screen and a whispered word that matched one of the on-screen words was played through headphones.[4] The subjects were asked to select the word spoken from among the options given. The subjects achieved 96% accuracy in the task; of the errors made, 36% were due to choosing the voicing cognate of the correct answer, 7% of the errors were manner-related confusions, and 56% were place-related confusions (p. 79, Figure 5.2). While the manner of a sound seems to be transferred best into whispered speech (i.e., it is easiest to identify), voicing confusion was less common than place confusion, and the overall error rate was quite low.

1.3 Glottal and laryngeal articulation of phonological voicing in normal and whispered speech

The glottis and the larynx are anatomically distinct. However, the use of the terms "glottal" and "laryngeal" are sometimes conflated, which renders the meaning of terms like "supraglottal" and "supralaryngeal" ambiguous. In discussions of phonation, it is particularly important to define these terms precisely. Barnes et al. (2005: 111) provide a clinical (surgical) definition, dividing the larynx into three distinct "compartments": supraglottis, glottis, and subglottis. The supraglottal laryngeal structures include the aryepiglottic folds, false vocal folds (vestibular folds), and the epiglottis. The glottis is limited to the (true) vocal folds and the structures that join it to the subglottis and supraglottis (the musculature and cartilage 10 mm above and 10 mm below the vocal folds). The subglottis extends from 10 mm below the vocal folds to the inferior end of the cricoid cartilage. I will use the term "supralaryngeal" to refer to the vocal tract above the vestibular folds, i.e., above the structures commonly associated with the larynx. I will use "glottal" and "laryngeal" in the sense of Esling (1984): "glottal" refers to the opening along the length of, and including the vocal folds themselves, whereas "laryngeal" refers to musculature

[4] One example: Lim played "dent" while showing "dent", "tent", "rent", "went", "bent", and "sent" on the screen.

and cartilage directly above and directly below the glottis (excluding the vocal folds themselves).[5]

The glottal and laryngeal differences between voiced and voiceless consonants during normal speech are well studied. Ladefoged (1971: 18) describes the phonatory setting for voicelessness as a state in which “the anterior portion of the glottis is so far apart that it cannot be set in vibration”. Abduction of the vocal folds, controlled primarily by the posterior cricoarytenoid muscle, is most often associated with the absence of vibration during voicelessness (Löfqvist and McGarr, 1987). However, Löfqvist et al. (1989) found higher levels of cricothyroid muscle activity for voiceless consonants than for their voiced congeners, suggesting that an increase in the longitudinal tension of the folds also inhibits vibration.

Whispered speech is sometimes called “nonvocal sound production”, though this definition seems insufficient for distinguishing whisper from mere voicelessness (Zemlin, 1998: 176). The glottal state is often characterized as “open” during whisper, which prevents the operation of the Bernoulli Effect and therefore inhibits vocal fold vibration, as happens in the voiceless setting. However, the glottis and larynx play a more active role in whisper than this simple characterization suggests. During normal speech, the arytenoid cartilages are held parallel with their medial surfaces contacting each other, allowing for vocal fold vibration and, therefore, phonetic voicing. During whisper, the arytenoid cartilages instead abduct slightly to produce an inverted, “Y-shaped” configuration, with the vocal folds touching along most of their length, except for a small gap (Zemlin, 1989). Impressionistically, the configuration has been variously described as “V-shaped”, “bowed”, and “slit” (Solomon et al., 1989).

Lindqvist-Gauffin (1969: 30) includes structures of the supraglottal larynx in his characterization of whisper, noting that the aryepiglottic sphincter above the glottis is contracted. Thus, the source of acoustic excitation during whispered sounds is turbulence, not vibration, at and/or directly above the glottis.

---

[5]Esling & Harris (2003), Lindqvist-Gauffin (1969), Zemlin (1998), Zeroual et al. (2005) also differentiate between the glottis and larynx. In some studies of whisper, e.g. Stathopoulus et al. (1991) and Mills (2009), the definition of these terms is less clear.

The (open) glottal configurations associated with respiration and whisper are said to be similar (Zemlin, 1998: 176). However, whisper produces a more turbulent, noisy airstream. Esling (1984) observes that, although they share some features, whisper and breathing are distinct in both glottal and laryngeal configuration. He attributes the differences to greater tension in the larynx and a narrower glottal opening during whisper. Esling and Harris (2003: 1051) further elaborate the difference between breath and whisper, stressing that it is not the configuration of the glottis, rather, "the shape of the epilaryngeal channel formed by the fronted and raised cuneiform cartilages at the 'elbow' of the aryepiglottic folds" that defines a whisper. It is presumably this supraglottal difference which serves to increase the degree of turbulence in the airstream during whisper. The greater turbulence in the larynx increases the amplitude of the noise, rendering whisper generally more audible than breathing.

It is still unclear whether glottal shape and size differ according to type of whisper, e.g. LEW vs. HEW. Monoson and Zemlin (1984) note that glottal configuration changed from a bow shape or inverted-V shape to a slit or inverted-Y shape as the speaker shifted from LEW to HEW. Laver (1980) and Greene (1980, as reported by Solomon et al., 1989) report an inverted-Y shape for both types of whisper, but note that the glottal gap narrows as noise intensity increases. Solomon et al. (1989), however, found no change in glottal shape or size based on whisper intensity: all vocal-fold configurations appeared to be straight slits or slits with a slight forward convergence near the vocal processes; the glottal openings observed were predominantly "medium"-sized.[6] While no systematic glottal changes were observed, the authors did uncover evidence of a laryngeal adjustment pattern between the two whisper types, wherein anterior supraglottalic constriction was higher in HEW than LEW. The authors also noted a "devoicing laryngeal gesture for /t/", which occurred in both normal and whispered speech modes (p. 172).[7]

In a laryngoscopic study, Zeroual et al. (2005) found no glottal or laryngeal differences that consistently distinguished voiced from voiceless consonants across both whispered and normal speech modes. While glottal aperture differences were observed between voiced and voiceless

[6]Solomon et al. used rough categories of glottal opening sizes (small, medium, and large), based on video recordings taken using fiberoptic endonasolaryngoscopy (p. 163).
[7]The authors mention this phenomenon in the discussion of findings, but fail to include any information about it in the results section, so the degree and specifics of this change are unknown.

cognates in normal speech, in whispered speech, those differences appeared to persist only in fricatives, not stops.

In his study of the effects of whisper on glottal behavior in both vowels and voiced/voiceless consonant pairs, Mills (2009) found results that contradict those of Zeroual et al. (2005). First, while speech mode did not affect the size of the glottal opening, more variability was observed in whispered consonants than normal consonants, indicating that glottal configuration (defined, in this case, solely by the size of the glottis) is more variable in whispered versus normal speech. Second, the voiced (adducted) versus voiceless (abducted) glottal size variation well-known to normal speech was maintained during whisper. When combined with the result of his previously-discussed perceptual experiment (listeners can discriminate voiced from voiceless whispered consonants even when duration is held constant), this suggests that acoustic cues may be transmitted by the different glottal gestures. Mills conjectures that these cues gain salience in the absence of physical voicing, though it is still unclear how a narrower glottis, in the absence of actual vibration, can by itself result in an acoustic cue similar to voicing.

1.4 Aerodynamics of whisper and voicing contrasts

The aerodynamic characteristics of sounds produced in normal and whispered speech modes are still debated. While it is relatively well-accepted that voiced and voiceless consonants (in normal speech mode) differ in their aerodynamic requirements (Ohala, 1983), it is not clear how these results relate to whispered speech in particular. Studies reviewed in this section indicate that whispered speech is slower and requires more air expenditure per syllable than normal speech (Schwartz, 1972; Stathopoulos, et al. 1991). Results regarding intraoral air pressure are equivocal: lower pressure has been reported in whispered versus normal speech (Murry and Brown, 1976; Schwartz, 1972; Stathopoulos et al., 1991), while others have found evidence of higher intraoral pressure in whisper (Klich, 1982; Murry and Brown, 1976), and still others find intraoral pressure to be equivalent during both modes (Weismer and Longstreth, 1980). More consistency is seen in studies of air flow during whisper, where it is generally shown to be greater than during normal speech (Schwartz, 1972; Stathopoulos et al., 1991). Voiced and voiceless stops also differ in oral air pressure and air flow, with voiceless stops predictably manifesting higher pressure and air flow than voiced stops, due to differences in glottal impedance. Some aerodynamic differences are limited to normal speech, but others extend to whispered speech as well (Weismer and Longstreth, 1980). Lastly, I will review Hoole et al.'s (1998) findings that aerodynamic conditions affect articulatory patterns in normal speech: ingressive air flow was shown to reduce or arrest forward movement of the tongue during production of velars and alveolars.

Schwartz's (1972) study of bilabial closure durations for normal and whispered /p/, /b/, and /m/ in English uncovered some ways that whispered consonants (and vowels) differ from their normally-spoken counterparts. First, in a production experiment involving 12 male speakers, Schwartz found that, at the phrasal level, whispering a sentence takes longer than normally uttering the same sentence. The syllables of the sentences significantly increased in duration by an average of 14 ms (p. 2025). Using a spirometer with six male speakers, Schwartz found that, on average, a whispered syllable consumed 45% more air than a normally-spoken syllable (p. 2026).

Schwartz also examined the intraoral pressure of /p/ in 16 females, to determine if the increased consumption of air could be attributed to increased expiratory effort; if so, Schwartz reasoned that the whispered /p/ should have a greater intraoral pressure than normal /p/. He found the reverse: whispered /p/ was consistently produced with less pressure than its normally-spoken counterpart, which would indicate that subjects actually decreased expiratory effort during the production of whispered speech.

Stathopoulos et al. (1991) found that a solid majority (8/10) of speakers used nearly identical lung volumes for normal and whispered speech (p. 765). However, the authors also found that whispered speech, when compared with normal speech, was associated with: (1) lower peak oral pressure, (2) higher average flow, (3) lower laryngeal resistance, (4) fewer syllables per breath, and (5) more air volume expended per syllable. This is consistent with the observations of Schwartz (1972).

Weismer and Longstreth (1980) studied peak intraoral air pressure and peak flow for /b/ and /p/ under both normal and whispered speech modes. Contrary to Stathopoulos et al. (1991) and

Schwartz (1972), they found no significant difference between air flow in whispered versus normal speech. Moreover, peak pressure failed to show a significant difference between whispered and normal speech. However, two significant results were found: (1) /p/ and /b/ differ in oral pressure in normal speech (/p/ > /b/), and (2) peak air flow was significantly higher for /p/ than for /b/ in both whispered and normal speech. The authors argue that distinct laryngeal gestures underlie and differentiate voiced and voiceless stop cognates in normal and in whispered speech.

Klich (1982), in contrast to the authors just discussed, observed increased intraoral pressure during whisper (versus normal speech). Also, contrary to Weismer and Longstreth, Klich reported higher intraoral pressure in the unvoiced consonant (/p/) than in the voiced consonant (/b/) during whispered speech alone; this difference was not observed in normal speech. Murry and Brown (1976) came to a different conclusion, finding that intraoral pressure during whisper varied by subject. For some subjects, whispered intraoral pressure was higher than intraoral pressure in normal speech; in other cases, a subject's normal speech manifested greater intraoral air pressure than whispered speech.

Sundberg et al. (2010) also studied the aerodynamics of whisper, but expanded their examination to include four types of whisper, produced by a trained speaker: hyperfunctional, hypofunctional, neutral, and postphonatory. The speaker produced the first three whisper types by consciously controlling his laryngeal opening (using a video monitor) so that hypofunctional whisper manifested the largest glottal opening, hyperfunctional manifested the smallest, and neutral whisper manifested a glottal opening between these extremes. Postphonatory whisper was defined as "whisper directly after a brief phonation" (p. 577). The authors focused on subglottal pressure instead of oral pressure. They also examined oral air flow. The authors found that hyperfunctional whisper (small glottis) was associated with less flow and higher subglottal pressure than hypofunctional whisper (large glottis). Neutral and postphonatory whisper values fell between these extremes. Although Sundberg et al. limited their comparison of subglottal pressure as a function of speech types to different varieties of whisper, similar alterations might occur between normal and whispered speech, though most likely at different levels.

Finally, Hoole et al. (1998) conducted an articulatory study focusing on the effect that varying aerodynamic conditions have on the supralaryngeal articulation of voiced consonants. For velar stops in particular, lingual articulatory loops have been reported during egressive speech: the tongue reaches its velar target and continues moving forward, then returns to its original position in a "looping" gesture (Houde, 1968; Perkell, 1969; Kent and Moll, 1972). Hoole et al. tested the possibility that these forward loops are caused by increased air pressure behind the tongue. Five phonetically-trained subjects produced VCV tokens featuring voiced velar and alveolar consonants /g/, /d/, and /n/ in pulmonic egressive and pulmonic ingressive flow modes. The authors reasoned that if a forward articulatory loop manifested itself even during ingressive speech, the loop could not be attributed to a posterior air pressure build-up. Under the ingressive flow condition, the velar tokens manifested forward lingual movement, but the loop was much attenuated. Moreover, tongue position for alveolars was more posterior during ingressive flow. The authors conclude that aerodynamic conditions directly influence the movement of the tongue during normal speech. However, aerodynamics is not entirely responsible for the observed forward loop, which still occurs for ingressive velars, albeit with a reduced amplitude of movement. This study suggests that during phonation modes associated with higher rates of transglottal flow (like whisper), the linguopalatal contact of consonants could be more anterior due to increased pressure or air flow behind the occlusion.

1.5 Supralaryngeal articulation of phonological voicing in normal and whispered speech

Supralaryngeal articulatory changes (e.g., durational and place of articulation) produce acoustic cues that may help listeners discriminate whispered consonants that vary in phonological voicing. Studies reviewed in this section show that whispered consonants are longer than consonants uttered in normal speech mode, and that whispered plosives and fricatives show significant length differences more often than whispered nasals (Schwartz, 1972; Parnell et al., 1977; Jovičić and Šarić, 2008; Higashikawa et al., 2003). While place of articulation (POA) changes in whisper have not been reported (Yoshioka, 2008), voiced fricatives have been shown to have a more anterior POA than their voiceless cognates (Fuchs et al., 2007; McLeod et al., 2003). Additionally, voiced fricatives may have a higher articulatory center of gravity and be articulated with greater amount of contact than voiceless fricatives (Fuchs et al., 2007). In whispered speech only, the voiced bilabial stop /b/ manifested higher peak opening velocity, peak closing velocity, and maximum lip separation than /p/ (Higashikawa et al., 2003). A study by McLeod (2006) examining the articulation of /n/ (an alveolar consonant with no phonologically voiceless cognate in English) shows a high degree of both intra- and inter-speaker variation in average COG and total palatal contact. McLeod argues that the articulation of /n/ requires less precision (versus a consonant like /s/) to achieve equivalent acoustic results. Similarly, alveolar fricatives have been found to require more articulatory precision and result in less palatal contact than alveolar stops: in the first case, the tongue is placed on the roof of the mouth; in the second, the tongue moves forcefully to a target "beyond" the palate, allowing the collision between tongue and palate to determine the exact placement and amount of contact (Mooshammer et al., 2003; Fuchs et al., 2006; Löfqvist and Gracco, 1997). Hyperarticulation, then, may be associated with less contact than hypo- or normal articulation. Lindblom's (1990) H&H (Hyper- and Hyposeech) theory asserts that a speech mode (such as whispered speech) can act as a speech perturbation that requires compensatory actions. Because compensatory actions cause individuals to slow their speech, a higher accuracy in target attainment is possible. This allows one to use the degree of variance observed as a litmus test for hypo- or hyperarticulation.

1.5.1 Acoustic and aerodynamic studies

Schwartz (1972) found whispered syllables to be longer than normal syllables. He hoped to determine if increased syllable duration was due to an increase in consonant duration, or if it was solely due to vowel duration as previously thought (Schwartz, 1968). Seven female speakers produced the phrase "VCV saw VCV with VCV", where the vowels were /i/ and /a/ and the consonants were /p/, /b/, and /m/, e.g., "apa saw apa with apa." Schwartz (1972) found significant effects for neither phrasal position nor vowel environment, but did find significant increases in duration for both whispered /p/ and whispered /b/. No significant duration differences were found for whispered /m/. The oral stops /p/ and /b/ had significantly longer closures than /m/ for both speech modes. According to Schwartz, this provides evidence for what I will call the "Respiratory Conservation hypothesis": to compensate for high air flow during whisper (conditioned by low glottal resistance), speakers "conserve air... by prolonging air-arresting articulatory gestures", e.g., the occlusion during a stop (p. 2025). Schwartz pitted this hypothesis against what I will term the "Intelligibility hypothesis": speakers slow their speech during whisper to increase intelligibility (Schwartz, 1968). Schwartz (1972) tested this by using various levels of masking noise played back to the speaker. Because the speech rate of his subjects did not slow in proportion to the level of masking noise introduced, he concluded that the slower speech during whisper was not compensation for a situation with low comprehensibility.

Parnell et al. (1977) followed Schwartz's production study, but focused on alveolar consonants /t, d, n, s, z/ instead of bilabials. The same vowels and carrier phrase were used. However, in this study, ten speakers of American English produced 20 sentences each (the five consonants in each of the two vowel environments with both normal and whispered phonation). The authors found that the mean length of whispered sentences was slightly higher than that of the normal sentences (1.8 seconds vs. 1.6 seconds, p. 612).[8]

As in Schwartz's study, no effect was observed for sentential position, so the data were pooled for the remainder of the analysis. Segmentation criteria for the whispered consonants were determined by (1) the absence of an F2 band during the stops and (2) the presence of turbulence around 8 kHz during the fricatives. The authors found that /t, /s/, and /z/ were significantly longer in whispered versus normal speech for both vowel environments. They found that /n/ was slighter longer in whispered mode, but not enough to be significant. Contrary to Schwartz, Parnell et al. (1977) found that whispered /d/ was significantly shorter than normal /d/. However, the authors acknowledged great variability in the spectographic patterns for both /n/ and /d/. Furthermore, the variation in the production of /d/ may have been due to several subjects who produced a tap instead of a stop, most often in whispered speech. This could account for the longer /d/ duration in normal speech. However, it was found that the voiceless cognates (/t/ and /s/) were longer than their voiced pairs (/d/ and /z/) under both normal and whispered conditions, which parallels Schwartz's findings for the bilabials.

Parnell et al. (1977) also concluded that both /a/ and /i/ were significantly longer during whispered speech versus normal speech. According to Schwartz's (1972) Respiratory Conservation hypothesis, speakers should decrease the duration of whispered vowels (which

[8]The authors did not specify whether this effect was significant.

exhibit a higher air flow), not increase it. In light of this, Parnell et. al. (1977) argued that the Intelligibility hypothesis (which Schwartz rejected) was more apt to explain this phenomenon.

Jovičić and Šarić (2008) undertook a broader acoustic analysis of whispered consonants. All 25 Serbian consonants were used (see Table 1). For the study, each consonant was positioned in an /aCa/ token inserted into a carrier phrase (in Serbian) three times: "*/aCa/ i /aCa/ su dva /aCa/*" (English: "/aCa/ and /aCa/ are two /aCa/"). Three male and three female speakers of Standard Serbian produced the sentence once with each consonant, under both normal and whispered voicing. They followed two different criteria for segmentation, taking into account the different acoustic patterns associated with whispered and normal speech: all normally-phonated speech was segmented according to the same criteria, whereas the whispered speech tokens were segmented differently by manner of articulation, using the cues available and appropriate to each.

Plosives	**Fricatives**	**Affricates**	**Nasals**	**Semivowels**
voiceless /p t k/ ***voiced*** /b d g/	***voiceless*** /f s š h/ ***voiced*** /z ž/	***voiceless*** /c ć č/ ***voiced*** /đ dž/	***voiceless*** /m n nj/	***trill*** /r/ ***laterals*** /l lj/ ***approximants*** /v j/

Table 1: Serbian consonants by manner of articulation

The authors found that whispered voiced plosives, fricatives, and affricates show a duration increase of 13.2%, 13.5%, and 19.1%; voiceless fricatives and affricates show an increase of 5.9% and 7%. Whispered nasals manifest a duration increase of 17.1% (p. 268). Voiceless plosives and nasals both showed small but non-significant results. They interpret this to mean that nasals and semivowels pattern similarly to the other voiced consonant. So, while voiceless consonants have greater durations than their voiced congeners in both whispered and normal speech, voiced consonants undergo greater lengthening in whispered speech than do voiceless consonants. Testing the duration with Pearson's correlation test, it was found that voiceless affricates have the greatest amount of variability in articulation (i.e., the least stable articulation), whereas semivowels exhibit the least variability (i.e., the most stable articulation).

Jovičić and Šarić also examined consonant duration by place of articulation (POA) and found that the increase in duration under whisper is significant for consonants at the bilabial, dental, alveolar, and palatal places, but not the labiodental or velar places. Consonants articulated in the palatal region saw the greatest durational increase during whispered speech (versus normal speech).

In addition to total consonant duration, Jovičić and Šarić examined VOT and the duration of affricate release. VOT for the whispered plosive tokens was defined as the time between the "strong burst of energy" and the "formant onset points of the vowel that follows" (p. 266-7). They found that the VOT duration for voiced plosives is significantly longer under whisper; however, no change was found for voiceless plosives between the two voicing conditions. Similarly, the affrication of voiced affricates increased in duration under whisper; however, no change was observed for voiceless affricates. The authors took this as evidence that voiced consonants exhibit a prolonged duration at both the subphonetic and phonetic levels during whisper.

Jovičić and Šarić interpret the results of prolonged duration as evidence of increased difficulty in articulating speech during whisper due to a greater precision required in the motor control of tongue movements. Furthermore, because voiced consonants have a greater time extension than unvoiced, voiced consonants are said to be more difficult to articulate properly than unvoiced consonants (at least during whisper). Likewise, the palatal consonants, which had the greatest duration increase by POA, are considered by the authors to be the hardest to articulate.

1.5.2 Lip and jaw kinematics studies

Higashikawa et al. (2003), examined the lip kinematics of /b/ and /p/ under normal and whispered speech. They did so by attaching reflective markers at the vermilion borders of the upper and lower lips of seven male speakers of American English. The participants were asked to either speak normally or whisper the phrase "my baba puppy" or "my papa puppy." Each speaker produced two blocks of 40 utterances. Each block was produced in its entirety in either normal or whispered speech. The order of presentation (whispered or normal block) was randomized.

These authors report four significant findings. First, in whispered speech, both the peak opening velocity and acceleration at the moment of peak opening velocity are significantly higher for /b/ than /p/. Second, maximum lip separation is greater overall in whispered than in voiced speech. Third, during whispered speech (but not normal speech), lip separation is greater for /b/ than /p/. During whispered speech only, peak closing velocity is greater for /b/ than for /p/. No significant differentiation of voiced / voiceless consonants in normal speech mode was reported.

The authors suggest that the rapid peak opening for /b/ would reduce the turbulence at the lips, altering the spectrum of the shortened burst. They posit that the increased lip separation of /b/ would further reduce turbulence, lessening the amplitudes of the high frequencies associated with the whispered production of /b/. We could then expect the release of /b/ to be quieter than that of /p/, with less subsequent frication. Because normal voiceless consonants tend to have a higher intensity relative to the following vowel than do their voiced counterparts, this change could help the listener distinguish phonemic voicing in the absence of phonetic voicing (Repp, 1979). The authors assert that, with VOT unavailable during whisper, the relative difference in intensity may be leveraged by speakers to accommodate listeners who might otherwise struggle to hear the difference between voiced and voiceless.

Contrary to Higashikawa et al., who found no differences in movement or displacement between voiced and voiceless cognates in normal speech, Sussman et al. (1973) found a number of differences. Sussman et. al examined peak velocity values of lower jaw closing, lower lip closing and “net lip opening” (lower jaw plus lower lip) of /p/ and /b/, finding a slower movement for voiceless /p/ versus voiced /b/ in lip closing and net lip opening; however, they observed a faster closing movement for /p/ versus /b/ when examining jaw movement alone. The first two (lip closing and net lip opening) coincide with Higashikawa et al.’s findings for whispered speech; the last (jaw closing) contradict what Higashikawa et al. reported in whispered speech.

Gracco (1994) found results similar to Sussman et al., observing that the peak jaw closing velocity of /p/ was greater than that of /b/ (in normal speech). Fujimara and Miller’s (1979) results for /d/ and /t/ were analogous to those of both Sussman et. al and Gracco: they reported that the peak jaw closing velocity of the voiceless cognate /t/ was greater than that of /d/ (also limited to normal speech). Both Gracco's and Fujimara and Miller's results contrast with Higashikawa et al.'s (2003) observations of closing velocity differences between voiced and voiceless consonants in whispered speech. Because Higashikawa et al.'s data on whisper conflicts directly with the other authors' data on normal speech just discussed, it appears that movement patterns in whispered speech are reversed from those in normal speech.

1.5.3 EPG studies

Recently, researchers have begun using electropalatography (EPG) to study phonological voicing contrasts among stops (McLeod et al., 2006; Fuchs et al., 2007; Yoshioka, 2008; Gibbon et al., 2007) and corresponding nasals (McLeod, 2006; Gibbon et al., 2007). Only one study has used EPG to study whispered supralaryngeal articulation (Yoshioka, 2008) and this study does not compare whispered and normal tokens directly.

There is reason to believe that voiced and voiceless cognates might be articulated in different locations. The Aerodynamic Voicing Constraint (AVC; Ohala, 1983) suggests that consonants with a more anterior place of articulation are more likely to be voiced than more posterior consonants. This is because a larger supraglaryngeal volume (associated with anterior consonants) allows the absolute transglottal pressure differential to remain positive for longer than a small supralaryngeal volume (associated with posterior consonants). Ohala (1983) uses the AVC to explain the cross-linguistic tendency for the greater prevalence of voicing contrasts in anterior versus posterior stops and also suggests that supralaryngeal adjustments can be made to prolong voicing, particularly the slight expansion of the cheeks for buccal obstruents like /b/. It may be possible to use the constraint to predict anteriorization of voiced consonants (with respect to their voiceless congeners). This difference may persist or even become enhanced during whisper, in the absence of phonatory cues to the distinction. Though nasal stops require an oral occlusion, they allow for a continuous passage of air through the nasal cavity. Thus, they may not require the same adjustment for whisper observed in oral stops, where oral and nasal flow both cease during the oral occlusive phase. For this reason, articulatory adjustments observed in whispered oral and nasal stops are unlikely to share the same aerodynamic motivation.

McLeod et al. (2006) examined EPG recordings of ten speakers' realizations of both word-initial and word-final /s/ and /z/ in five vowel contexts (the number of utterances was not reported) in normal speech conditions. They found that voiced fricatives were significantly more anterior (as judged by contact across the front two rows) than voiceless fricatives. No significant differences in center of gravity (COG; as defined by Hardcastle et al., 1991) or total palatal contact were observed.

Fuchs et al. (2007) used EPG to examine the articulatory realization of voicing contrasts in German alveolar and postalveolar fricatives produced in normal speech conditions. Seven speakers (four male and three female) produced 19 tokens ten times each in randomized order, using the phrase *Sage* ___ *bitte* (English: "Say ___ please"). The fricatives occurred in word-intial, -medial, and -final positions. The authors found that the most reliable tokens for studying the production of true voicing contrasts (the tokens in which the voicing was phonetically-realized as well as phonologically-specified) were those where the fricatives occurred word-medially. Articulatory measures included the percent of contact over the entire palate (PC), the percent of anterior contact (ANT), and the COG averaged over the duration of the fricative. However, Fuchs et al. (2007) used a different measure of anteriority than McLeod et al. (2006), basing their calculation on the first four (rather than two) rows of the palate. The authors found that voiced fricatives had significantly higher ANT, COG, and PC values. The largest difference was found in the ANT values of the voiced and voiceless alveolars and postalveolars. The COG measures patterned with the ANT measures, but were slightly less pronounced. The PC values showed a small but significant change.

Yoshioka (2008) also used EPG to study voicing contrasts in fricatives, but focused on /s/ versus /z/ articulation in whispered speech only. Yoshioka examined the speech of a single Japanese-speaking female. Fifty randomly-ordered utterances that contained /s/ or /z/ were produced. Peak maximum contact was found to be greater in /z/ than /s/. The author found that the contact pattern for /z/ was more stable (i.e., it showed less variability over the duration of the fricative) than that of /s/.

Gibbon et al. (2007), in their EPG study of normally-spoken alveolar stop consonants, found no difference in variability between /t/, /d/, and /n/ in normal speech in either high or low vowel contexts. Furthermore, no difference in percent contact or shape[9] was observed between /t/ and /d/.

McLeod (2006) studied /n/ production in seven adult speakers of Australian English (four males and three females) using EPG. The author observed that productions of /n/ were highly varied and found a large degree of inter- and intra-speaker variability in both measures of average COG and average total palate contact, as well as variation in contact pattern. While most speakers had a fairly symmetrical (right-left) contact pattern of lateral bracing in the palatal and velar regions, two speakers (one male and one female) exhibited lateral bracing that favored the right side along the teeth. Most speakers exhibited an alveolar closure within the first three (anterior) rows of the palate. McLeod felt the substantial variability within and between subjects could be due to differences in anatomy, an early mastery of the phoneme (which fails to be subsequently fine-tuned by the child), or a greater tolerance of variability in /n/ production leading to the same acoustic result. According to McLeod (2006: 103), "we hear /n/ as /n/ across a wider range of possible tongue strictures than we would tolerate for, for example, /s/. Thus, the exact place of articulation may not be as crucial."

A difference in the accuracy of placement required for different manners of articulation (stops versus fricatives) was also observed by Fuchs et al. (2006), who found using EPG and EMMA that, in addition to the different aerodynamic constraints for fricatives and stops, different structural needs (a more complex pattern of tongue bracing) and movement patterns occur in the two. According to Fuchs et al., fricatives require a careful placing of the tongue on the lateral margins of palate, whereas in a stop consonant, the speaker simply aims "through" the palate and allows the tongue's collision with the palate to arrest movement (Mooshammer et al., 2003; Fuchs et al., 2006). The same has been said to occur for bilabial stops (Löfqvist and Gracco, 1997, as reported by Fuchs et al., 2006). Fuchs et al. furthermore found that this affected the amount of contact: stops tended to have more contact than fricatives (even when normalizing for

[9]The most common shape found was with an anterior constriction present and a posterior constriction absent.

the channel opening required of fricatives). Thus, the more careful articulation was associated with a decreased amount of contact between the tongue and palate.

1.5.4 Hyper- and Hypoarticulation

Finally, we turn from EPG results to a theoretical discussion of the articulatory phenomena of hyperarticulation and hypoarticulation. One foundational work on the topic is Lindblom's H&H (Hyper- and Hyposeech or Hyper- and Hypoarticulation) theory; this will be used to interpret differences in variance (and possibly, as discussed above, amount of contact) that may be uncovered using EPG.

Lindblom (1990: 403) based his theory on the observation that normal speech is characterized by a "lack of invariance". In other words, speech varies: speakers exhibit perfect preciseness in neither articulation nor acoustic output. This variance is caused by competing demands. On one hand, the speech system's natural tendency is to use the lowest-cost movements possible: Lindblom called this "economy" (p. 404).[10] On the other hand, speakers must ensure their speech exhibits "sufficient contrast" for intelligibility; that is, the signal must be precise enough to be "sufficient for lexical access" (p. 405). This requires adjustments constantly be made to keep the articulators close enough to the target to be discriminable; the ability of the motor system to change its movement if it is too far from its target is called "plasticity" (p. 404).

Internal and external factors dictate whether economy or plasticity is the predominant mode of operation at any moment. In the H&H theory, described by the author as "a deliberate simplification", speech is subject to two categories of controlling factors: system constraints and output constraints (Lindblom, 1990: 418-9). System constraints are limitations imposed by the physical articulators and internal representations of the production of speech sounds. These limitations might include the length of the tongue, the depth of the palatal vault, and the optimal physical realization of a particular consonant. System constraints can be thought of as those constraints that are always present. When system constraints are the predominating influence, the movement "tends to default to some low-cost form of behavior"; economy is the predominant mode in this situation (p. 404).

[10] All biological motor systems are said to be subject to this tendency.

Output constraints, on the other hand, are "add-ons" which arise due to (1) innate knowledge of the listener's ability to attend to the speaker's signal and (2) other external/unusual constraints placed on the speaker's articulators. These external/unusual constraints might include "perturbations" such as bite blocks, loud rooms, listeners with hearing deficiencies, and different speech modes. Any situation which causes the speaker to alter his or her speech pattern in a way that "is goal-directed and involves articulatory manoeuvres that are compensatory" may be considered a perturbation (Lindblom, 1990: 423). In these situations, speech production is highly "output-oriented" and plasticity is utilized (p. 404).

In H&H, system and output constraints relate directly to precision of articulation. Lindblom, (1990: 417) credits the output-oriented mode with producing "more accurate target attainment" and states that, in the system-oriented mode, "target attainment is less efficient". He further states that "[w]hen output constraints dominate, we expect to see hyperforms, whereas with system constraints dominating, hypospeech will be observed" (p. 418). Because different phenomena are sometimes denoted by the terms hyperarticulation and hypoarticulation, it is important to clarify this. In H&H, "overarticulation" (hyperspeech, hyperarticuation) is "speak[-ing] as clearly as possible" (p. 429).

Whisper, as an alternate speech mode, conforms to Lindblom's definition of an external or unusual constraint (output constraint) placed upon a speaker due to its different aerodynamic environment and (at least partially) degraded signal. Thus, whisper arguably acts as a perturbation; in Lindblom's (1990: 422) words, this is a "natural bite block". According to Lindblom, natural bite blocks require a compensatory gesture, as they are output-oriented. This means that, in whisper, we expect to see hyperarticulation.

What does hyperarticulation look like and how can we tell if it's occurring? Most obviously, variability in the speech signal should decrease if, as Lindblom states, the output-oriented mode increases the accuracy of the articulator's movement toward the target. Also discussed by Lindblom, but less obvious, is the presence of "undershoot", in which "the movement toward the...target is reduced" (Lindblom, 1990: 414). When individuals were given "explicit instruction to ‘overarticulate’, and ‘to speak as clearly as possible”, Lindblom found that “[a]ll speakers showed undershoot” (p 429). In vowels, this manifests as reduced forms; in alveolar consonants, we could expect to observe tapping or even instances of non-occlusion. However, because this was found to vary greatly by speaker and speech style, the production of full forms cannot serve as proof that hyperarticulation isn’t occurring.

While it might first seem puzzling that hyperarticluation/overarticulation would result in reduced forms or lack of contact, Lindblom (1990: 415) explains this as the effect of a speech system balancing compensation with its desire to “operate so as to minimize ‘articulatory effort’ (peak velocity)”. In other words, the speech system slows down in order to reduce the “biomechanical effort” used (Nelson 1983, in Lindblom 1990: 414). If coarticulation occurs (as it does in all speech), it can cause the target to not be reached (undershoot), as it moves on to the next speech sound before the vowel or consonant in question is fully-realized. In this way, the tongue’s aim is more exact and less variable, but the amount of contact might also be decreased.

To further discuss the relationship between amount of contact and preciseness of articulation, we return to Fuchs et al.’s (2006) previously-discussed EPG study of German alveolar obstruents. While it would be incorrect to categorically state that some manners of articulation manifest hyperarticulation and others manifest hypoarticulation, parallels can still be seen between the fricatives and stops. The fricatives, like hyperarticulated consonants, had careful placement on the palate and relatively low speeds. A substantial deceleration was seen as the tongue approached the palate. The stops, conversely, were loosely articulated (have a target above / through the palate), a much lower degree of deceleration was seen as the tongue neared the palate, and they exhibited much greater speeds. It is not a stretch, then, to posit that hyperarticulated consonants would show decreased surface contact (as fricatives did), even when there wasn’t undershoot, whereas hypoarticulated consonants (like stops) would produce occlusions with more contact.

This means that EPG, which can be used to measure both variance and contact, is a tool well-suited to study hypo- and hyperarticulation. If one compares an experimental population of data to “normal” speech and that experimental population shows increased variance, especially combined with a greater amount of contact, this would indicate that hypoarticulation is occurring. However, if variance and/or contact decreases, this serves as substantial evidence that the subject is hyperarticulating to compensate for an unusual speech condition (with whisper acting as an output constraint).

1.6 Research hypotheses

As reviewed in the previous sections, whispered speech differs in many respects from normal speech. This includes how speakers articulate phonemically voiced and voiceless consonants during whisper, when vocal fold vibration is impossible. Generally, while some differences between normal and whispered speech can be attributed to differing glottal and laryngeal configurations, others cannot. For example, whisper is commonly typified by higher air flow and higher subglottal pressure (Schwartz, 1972; Stathopoulos, 1991; Sundberg et al., 2010). On the other hand, consonant and vowel formants are altered during whisper, possibly indicating some kind of supralaryngeal adjustment.

If we consider how voiced and voiceless cognates are articulated in normal and whispered speech, we find other differences. In the absence of phonetic voicing, supralaryngeal articulations may play a role in transmitting phonemic voicing status. In studies of normal speech, some authors find no differences in the movement of supralaryngeal articulators between voiced and voiceless consonants (Gibbon et al., 2007 for normal speech; Higashikawa et al., 2003 for whispered speech). Others find differences in POA for normal speech (McLeod et al., 2006; Fuchs et al., 2007) or articulator speed (Sussman et al., 1973; Fujimara and Miller, 1979; Gracco, 1994). In whispered speech, supralaryngeal differentiation of phonemic voicing counterparts is more consistently reported: all studies reviewed showed some evidence of supralaryngeal differences in voiced and voiceless cognates, such as increased peak maximum contact in phonemically voiced fricatives (Yoshioka, 2008) and increased opening and closing speed of phonemically voiced bilabials (Higashikawa et al., 2003).

The present study contemplates a direct comparison of voiced and voiceless stops in whispered and normal speech using EPG. It will help determine whether supralaryngeal changes in duration, POA, and degree of contact can differentiate voiced from voiceless cognates in normal and whispered speech, and will assess the variability in each of these measures. By considering both normal and whispered speech, it will be possible to inquire how supralaryngeal voicing differences in normal speech are enacted in whispered speech.

Answers to these questions may help explain why whispered and normal speech differ. Do speakers articulate differently when whispering in order to reliably communicate their message in spite of a degraded (whispered) acoustic signal (i.e., the Intelligibility hypothesis rejected by Schwartz, 1972, but supported by Parnell et al., 1977)? If not, are speakers conserving air in order to extend the amount of time they can speak between breaths (the Respiratory Conservation hypothesis favored by Schwartz, 1972)? A final possibility is that changes observed are merely passive movements caused by the different aerodynamic conditions associated with whisper.

It is probable that some supralaryngeal differences in whispered speech are caused by the higher air flow and (possibly) higher intraoral pressure that characterize whispered stops. However, any articulatory changes (in whisper versus normal speech) that cannot be explained by the aerodynamic conditions alone must have a different underlying cause. In the absence of voicing, it is possible that supralaryngeal articulations (duration, POA, etc.) account for the listener's ability to distinguish phonemically voiced from voiceless consonants during whisper. Therefore, any supralaryngeal differences observed between phonemically voiced and voiceless consonants during whisper are likely to indicate articulatory changes intended to accommodate the listener, i.e., by making the contrast easier to identify.

In the experiment that follows, I will consider whether speakers alter their articulation of phonemically voiced and voiceless consonants to compensate for the lack of phonetic voicing as a discriminatory cue. Consonant duration, POA, and degree of contact are all factors that speakers might use to compensate for the absence of vocal fold vibration (typical of whisper). Each of these factors, and the research hypotheses associated with them, will be discussed in the sections that follow.

1.6.1 Duration

Many researchers have studied durational differences between normal and whispered speech on the one hand, and phonemically voiced versus voiceless consonants on the other. They agree that whispered consonants are longer than normally-spoken consonants. Furthermore, whispered consonants at certain places of articulation tend to lengthen more than others, though researchers do not agree which places are those most affected (Schwartz, 1972; Parnell et al., 1977; Jovičić and Šarić, 2008). Specifically, normally-uttered voiceless stops tend to have a longer occlusion than voiced stops (Mills, 2003). If a difference between phonemically voiced and voiceless consonants in normal speech is exaggerated during whisper, the motive may be perceptual. In other words, the speaker may be exaggerating a supralaryngeal (in this case, durational) distinction in the absence of a glottal/laryngeal one. If a difference exists during whisper, but in no greater magnitude than during normal speech, it may be that the duration of the occlusion is merely part of the motor program associated with the individual consonants.

1.6.2 Place of articulation (POA)

If we observe a more anterior POA in whispered versus normal alveolar stops, the reason could be aerodynamic. Because whispered speech and voiceless segments are consistently associated with higher air flow (due to decreased laryngeal resistance), whispered and/or voiceless stops may manifest a more anterior articulation than normal and/or voiced stops. This would support Hoole's (1998) findings that egressive flow (accompanied by higher back pressure) results in a more anterior tongue position. It is also possible that subtle differences in place of articulation are associated with normally-spoken /t, d, n/, independent of their aerodynamics, and that these articulatory cues are accentuated by speakers under the degraded acoustic conditions associated with whisper.

1.6.3 Amount of contact

The amount of linguopalatal contact, as measured using EPG, corresponds roughly to the vertical placement of the tongue and thus the degree of constriction between the tongue and hard palate (Fontdevila et al., 1994).

Because the evidence of intraoral pressure differences in whispered versus normal speech is equivocal, a difference in surface area contact cannot be conclusively associated with the aerodynamics of the two speech modes (Murry and Brown, 1976; Schwartz, 1972; Stathopoulos et al., 1991; Klich, 1982; Weismer and Longstreth, 1980). Rather, if EPG differences are found between normal and whispered speech, there must be an alternate reason for the changes observed; one possibility is hyper- or hypoarticulation. If whispered speech is more difficult to articulate than normal speech, and the speaker compensatorily hyperarticulates, his or her speech should manifest lower tongue velocity and less contact (Fuchs et al., 2006; Lindblom, 1990; Mooshammer et al., 2003).

The question of intraoral pressure by consonant, however, is fairly well agreed upon: voiceless consonants (like /t/) have been shown to have higher intraoral pressure than voiced (like /d/) in both normal and whispered speech (Brown et al., 1973; Klich, 1982; Weismer and Longstreth, 1980). Nasals like /n/ exhibit the least pressure due to velopharyngeal opening (Brown et al. 1973). If the surface area contact percentage data increases as oral pressure increases, it is likely that amount of contact is related to intraoral pressure.

1.6.4 Variability

The stability or instability of an articulation corresponds on the degree of articulatory precision exercised in the realization of a given phoneme (Fuchs et al., 2006; Lindblom, 1990). If greater variability is observed in whispered speech, it might suggest that whisper is hypoarticulated (Lindblom, 1990). However, if less variability is observed in whispered consonants, it would point to a greater degree of precision in articulation, i.e., hyperarticulation.

CHAPTER 2: METHOD

2.1 Instrumentation

Data were recorded using the Articulate Instruments (AI) software interface (v. 1.17; Articulate Instruments, Musselburgh, UK). Each speaker was recorded while wearing a thin acrylic electropalate manufactured by Incidental (Newbury, Berkshire, UK). Each electropalate was designed using a unique superior maxillary model cast by a board-certified orthodontist. Each electropalate has 62 silver electrodes arrayed across eight rows proceeding from front to back; the anterior-most row had six electrodes and the rest had eight (see the bottom of Figures 1 and 2 for examples of the layout). The electropalate was secured to the roof of the speaker's mouth with wire clasps mounted on the palate; these fit snugly around the 1st or 2nd molar, depending on the speaker's dentition. The palate was connected to the WinEPG multiplexer, to the WinEPG Palate Scanner EPG3.V2, then to the WinEPG EPG3 Serial Interface SPI V2.0 (SPI). Whenever the tongue touches an electrode, an electrical signal is sent through this chain. Given that the position of each electrode is known, the electrode signals are ultimately interpreted as indicating whether a specific region of the palate is in contact with the tongue (Baken, 1987; Hardcastle, 1972). The sampling rate of the scanner is 100 Hz.

Speakers wore a head-mounted AKG C520 cardioid microphone (Harman International, Stamford, CT). The audio signal was passed to a Grace m101 pre-amplifier (Grace Design, Boulder, CO) and then to one of two line inputs of the SPI.

From the SPI, the audio signals were passed to an AudioFire2 (Echo Digital Audio, Carpinteria, CA) IEEE 1394 serial bus interface for isochronous real-time data transfer to an HP xw4400 Workstation running Microsoft Windows XP (Version 2002, Service Pack 3). The EPG signal was passed from the SPI to the same computer via USB. Synchronization of the audio and EPG signals is a function of the WinEPG hardware and Articulate Instruments 1.17 software installed on the machine (Articulate Instruments Ltd., 2008b). Audio was digitally sampled at a rate of 22050 Hz.

2.2 Speakers

Four male speakers of American English between the ages of 22 and 34 participated in the study. Two were from the West Coast of the United States (California, Washington), one speaker was from Utah, and the fourth was from North Carolina.

The curvatures of speakers' EPG palates were calculated using the method outlined in Brunner et al. (2009). In Brunner et al.'s measurements, there are flat, medium, and domed palates: flat palates have measurements greater than 1.9 cm^{-1}, domed palates are those less than 1.6 cm^{-1}, and medium palates fall between. Speaker 1's (S1) palate curvature is 1.76 cm^{-1}, Speaker 2's (S2) is 1.57 cm^{-1}, Speaker 3's (S3) is 1.68 cm^{-1}, and Speaker 4's (S4) was 1.40 cm^{-1}. So, using Brunner et al.'s classification, S1 and S3 have medium palates, and S2 and S4 have domed palates.

Some studies indicate that flat palatal shapes are associated with lower variability of articulation and domed palatal shapes are associated with a higher variability. Because of this, it may be more difficult to find positive results with the currents speakers' data than it would be with a group of speakers with flat palates. A positive result in the measures is, therefore, a strong indicator of the presence of a genuine effect. This could actually be considered a positive in a study with relatively few speakers.

2.3 Materials

VCV tokens manifesting alveolar consonants /t, d, n/ in vowel environments of /i, a, u/ were used. The first and second vowels were identical. The tokens had stress on the first vowel and were pronounced in their full, unreduced forms. The tokens were produced by the subjects in both normally-spoken and whispered forms for both experiments. The test items (along with distractors) are listed in the Appendix.

Using a variety of vowel environments was desirable because it has been noted that "the contact pattern of a consonant will be altered... as the vowel environment changes and as jaw position alters the distance and orientation of the tongue and palate" (Baken and Orlikoff, 2000: 530, citing Butcher, 1989; Stone et al., 1992; Recasens et al., 1995). The contact pattern is subject to "significant random variation", as well (Baken and Orlikoff, 2000: 530). This makes it necessary to collect a relatively large number of repetitions. After reviewing numerous EPG studies of disordered speech, McAuliffe and Ward (2006: 199) recommend five repetitions of each target consonant, “to ensure that a representative picture of a participant's tongue-palate contact is gained”. In the current study, ten repetitions of each token will be produced by each speaker to increase the accuracy and robustness of the results.

2.4 Production task

In the production task, stimuli were presented in randomized order to the subjects on a screen preceded by a “w” or “n”. Cued by these initials (representing **w**hisper and **n**ormal), the researcher verbally prompted the speaker by saying the word “normal” in a regular conversational volume (when “n” proceeded the phrase) or by saying the word “whisper” in soft whisper[11] (when “w” proceeded the phrase). The subjects were instructed to use the speech mode requested and to try to approximate the volume that the researcher used.

Speakers uttered the randomized normally-spoken and whispered tokens interspersed with an equal number of parallel VCV distractors using the bilabial consonants /b, p, m/. They read tokens in the carrier phrase “Say *X* again” and completed ten repetitions of the entire list.

Ten repetitions of 36 tokens (including distractors) were produced by each subject, meaning a total of 360 utterances per speaker were produced (6 consonants x 3 vowels X x phonation modes x 10 repetitions). Therefore, there were a total of 1,440 tokens total for all utterances in all speakers. Because half of the recorded tokens were fillers, a total of 720 tokens will be analyzed.

2.5 Post-processing

In order to access a more flexible environment for multi-channel annotation, data were converted to a Matlab structure. The binary EPG data recorded by the AI software were converted to text files using code adapted from *loadepg.m* (Nguyen, 1996). The final Matlab structure combined the synchronized audio and EPG data. A customized graphical user interface allowed for visualization and annotation of both signals in tandem (Shosted, accepted). After completing a

[11]The “soft whisper” is the same as the LEW, as defined in Section 1.1.

rough annotation (see annotation protocol below) of the consonant in each token, a customized Matlab script was used to realign the boundaries to the first and last frames manifesting total horizontal occlusion. The script detected the presence of complete horizontal occlusion across the palate as defined by contiguous contact across up to three rows of EPG sensors (Cho and Keating, 2009). Examples of the acoustic waveforms and signals can be observed in the top and middle panels of Figures 1 and 2, with the occlusive portion demarcated between two black vertical lines; the associated raw palatograms can be seen at the bottom of each figure. In Figure 1, the break of the occlusion is visible at 1.73 seconds. Absence of occlusion was noted when it occurred and is treated statistically in Section 3.1.

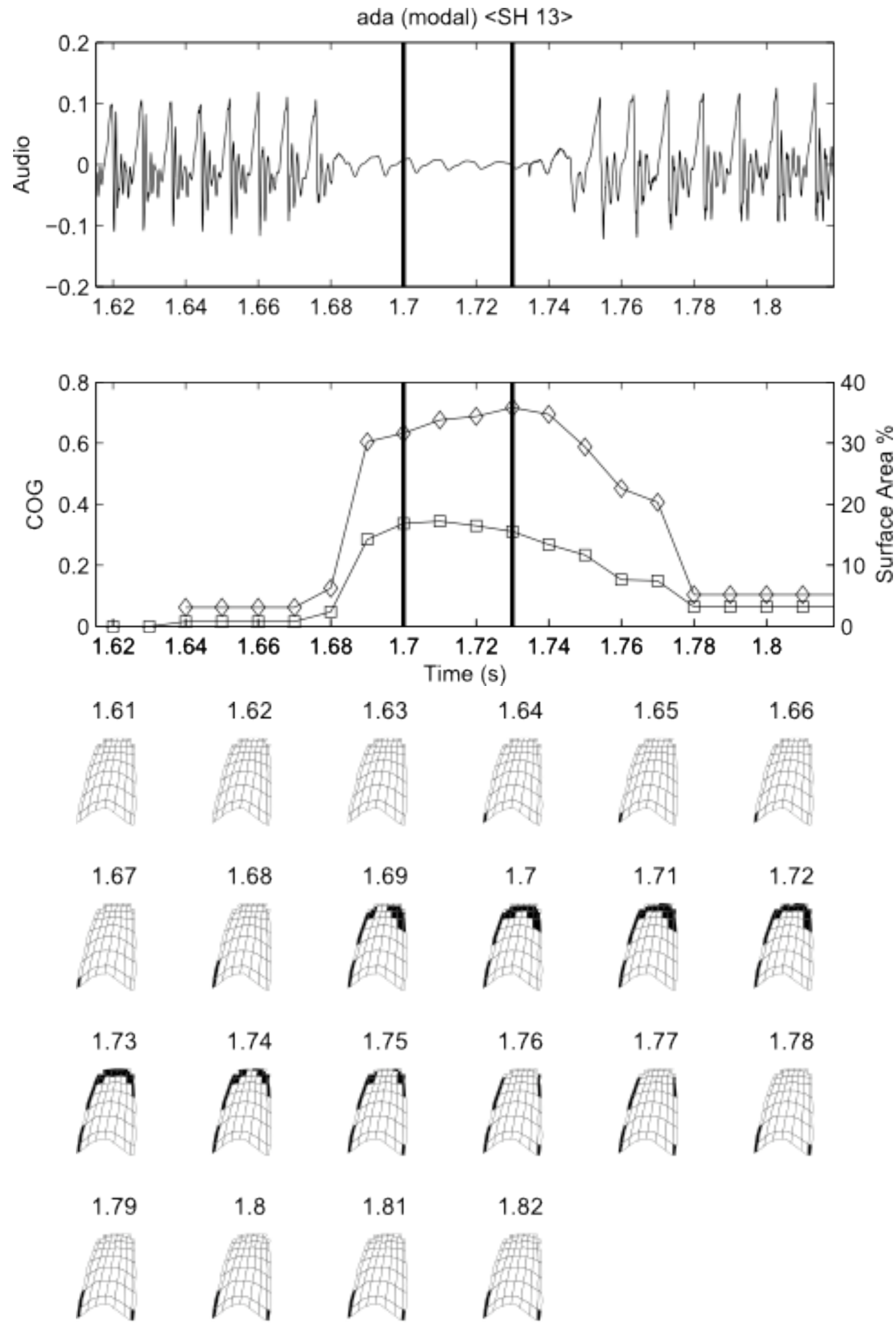

Figure 1: Example of a typical annotation of a normal /ada/ token with acoustic waveform (top), time-aligned signal (middle), and corresponding palate series (below). In the annotation, the boundaries of the rough (manual) annotation correspond to the edges of the acoustic waveform. The black bars show the beginning and end of the occlusive phase, as determined automatically by the Matlab script. Time-aligned surface area (squares) and COG (diamonds) signals are shown. The number above each palate (bottom) corresponds to the time in the acoustic waveform. Note the full occlusion that begins at 1.7 seconds and ends at 1.73 seconds.

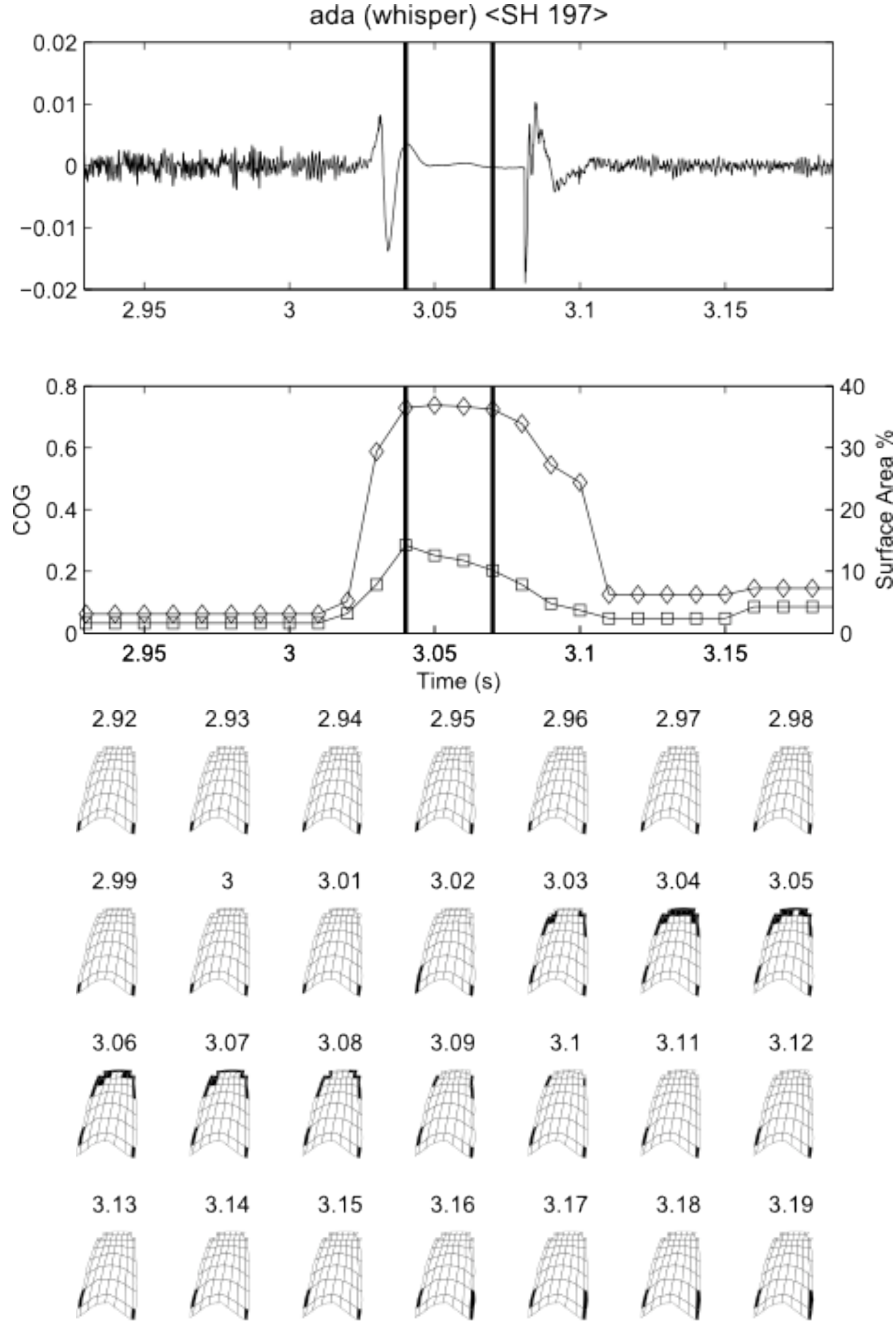

Figure 2: Example of a typical annotation of a whispered /ada/ token with acoustic waveform (top), time-aligned signal (middle) and corresponding palate series (bottom). In the annotation, the boundaries of the rough (manual) annotation correspond to the edges of the acoustic waveform. The black bars show the beginning and end of the occlusive phase, as determined automatically by the Matlab script. Time-aligned surface area (squares) and COG (diamonds) signals are shown. The number above each palate (bottom) corresponds to the time in the acoustic waveform. Note the full occlusion that begins at 3.04 seconds and ends at 3.07 seconds.

In addition, the following measures were calculated for the occlusive phase of each token:

- **Duration**: length of occlusion, as defined above.
- **Surface Area Contact Percentage (SACP)**: designed to quantify the *surface area* contacted (Shosted, in prep.) as opposed to the more frequently-used *percent of sensors* contacted (e.g. Fontdevila et al., 1994, among many others).
- **Center of Gravity: (COG):** designed to quantify the region where most of the linguopalatal contact is occurring by assigning progressively greater weight to more anterior rows of electrodes. This was measured for each EPG frame using the formula discussed below.
- **Maximum Center of Gravity (Maximum COG)**: This is the highest measure of COG observed during the occlusive phase.
- **Average Center of Gravity (Average COG)**: This is an average of all COG measurements observed over the duration of the occlusion.

The Duration was measured by subtracting the time associated with the first frame where a full occlusion occurred from the last frame where a full occlusion occurred.

The SACP is based on the unique dimensions of each subject's EPG unit. The 2D positions of the sensors along the length and width of the palate were measured after imaging each palate on a flat-bed scanner (Brunner et al., 2009). The area around each sensor was calculated by determining the midpoints between neighboring sensors and then computing a four-sided polygon, with a known surface area, for each sensor (Shosted, in prep.; see Figure 3). When a sensor is activated, it is assumed that the surface area of its associated polygon is in contact with the tongue.[12] After the total contact surface area is calculated for an individual EPG frame, this value is divided by the total surface area of the palate (determined by summing all surface area measurements for all sensors) to arrive at a percentage.

[12]Because the EPG palates are customarily manufactured such that posterior sensors account for more surface area than anterior sensors, traditional measures of contact percentage perhaps misleadingly treat each sensor as representing an equal portion of the hard palate. While this normalized contact percentage may correlate acceptably with the actual surface area contacted for some subjects or at some places of articulation, depending on the size and shape of the subject's oral cavity, this more traditional method introduces some noise into the resulting measurements.

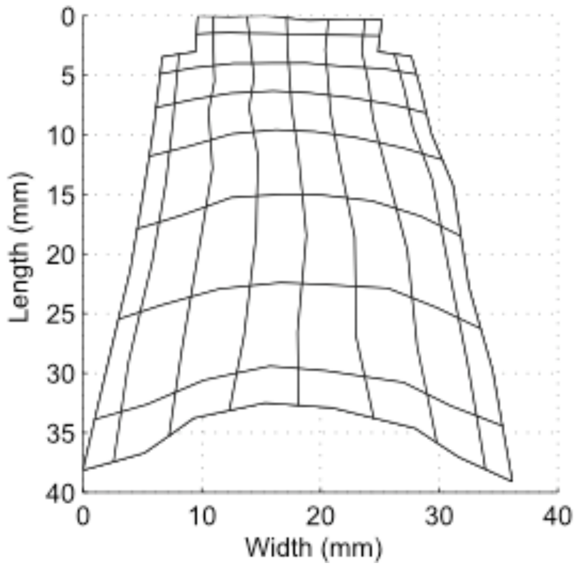

Figure 3: 2D representation of Speaker 1's palate. Positions of each electrode were measured after the palate was imaged with a flat-bed scanner (Brunner et al., 2009). Each four-sided polygons surrounds an individual sensor; together these are used to calculate the SACP.

The Maximum COG and Average COG were calculated using the center of gravity measure described in Articulate Instruments Ltd. (2008b) and Hardcastle, Gibbon, and Nicolaidis (1991). The calculation is based the following equation:

$$1 - \frac{\sum_{m=1}^{8}(m-0.5)R_m}{8\sum_{m=1}^{8} R_m}$$

where $R_m = \sum_{n=1}^{8} c_{m,n}$ is the sum of contacts in Row is m and $c_{m,n}$ is the contact value in the mth row and nth column (either 0 or 1). In the present formulation, center of gravity values range from a maximum of 0.9375 if only the anterior-most row electrodes are lit to a minimum of 0.0625 if the only the posterior-most electrodes are lit. A higher center of gravity value is generally associated with a more anterior place of articulation since the more anterior rows of electrodes are weighted more heavily in the calculation.

Both the maximum and average of a measure were calculated for each COG; for the latter, the measures were averaged over the time of total occlusion (Duration) for each token.[13] The SACP was also averaged over the Duration of the closure. Therefore, a total of four measures were calculated for each token: Duration, Maximum COG, Average COG, and SACP.

2.6 Annotation Protocol

A rough annotation was needed for all tokens before the boundary realignment was performed in Matlab. The phonetic initiation and termination of [d], [t], and [n] were defined in strictly articulatory terms based exclusively on EPG data. The following protocol was adhered to while using the custom-designed Matlab GUI: (1) Annotate just before the first point at which total linguopalatal occlusion occurs; (2) Annotate just after the last point at which total linguopalatal occlusion is recorded; (3) If linguopalatal occlusion does not occur (i.e., there is no fully-occluded [n], [d], or [t]), tally it a as ``non-occlusion." Frequency data for all tokens (normal and whispered) are given in Table 2. The outcomes of a two typical annotations are shown in Figures 1 and 2.

Speaker	Tokens manifesting occlusion (Norm/Whisp)	Tokens without occlusion (Norm/Whisp)	Percent (%) of tokens with occlusion
S1	46 (31/15)	44 (14/30)	69% normal > 33% whispered
S2	46 (21/25)	44 (24/20)	47% normal < 60% whispered
S3	85 (43/42)	5 (2/3)	96% normal ~ 93% whispered
S4	88 (44/44)	2 (1/1)	98% normal ~ 98% whispered

Table 2: Number of occluded and non-occluded tokens for each speaker. Each speaker produced 90 tokens (45 normal, 45 whispered). Tokens without occlusions were not included in analyses of occlusive characteristics.

2.7 Statistical analysis

Statistical analyses were conducted in R 2.8.1 and all function names correspond to this software. Contingency tables based on the presence/absence of occlusion were created using the

[13]The measure was taken for each EPG frame in which total occlusion occurred. These measures were added together. The sum was divided by the total number of frames to get an "average" COG.

xtabs function and the results were then submitted to Pearson's chi-square test for Count data (*chisq.test*). Tokens which failed to manifest a complete occlusion were excluded from further statistical tests, i.e., those dealing with the characteristics of the occlusion itself. Duration, Maximum Center of Gravity (Maximum COG), Average Center of Gravity (Average COG), and SACP during occlusion were submitted to analysis of variance. For each speaker, the center of gravity and surface area measures (maximum and average) for the ten repetitions of each token were averaged in a repeated measures design using the *recast* function in R’s *reshape* package. Mean and standard deviation values were incorporated in linear mixed effects (LME) models using the *lme* function in R's *nlme* package. Consonant (/t/, /d/ /n/), Vowel (/a/, /i/, /u/), and Speech Mode (normal, whisper) were included as fixed factors and Speaker was included as a random factor (Baayen, 2008). Individual LME models were designed for each speaker, with fixed effects as above (Consonant, Vowel, Speech Mode) and Speaker as random effect.

CHAPTER 3: RESULTS

3.1 Frequency data

13.2% (95/720) of the analyzed tokens lacked a complete occlusion: this included 11.4% (41/360) of the normal tokens and 15% (54/360) of the whispered tokens. A chi-squared test failed to show any significant association between absence of occlusion and speech mode [$\chi^2(1,720) = 1.75$, $p > 0.05$].

The same association was then tested for each speaker; relevant frequency data is provided in Table 2 above. For three of the four speakers, chi-squared tests uncovered no significant association of occlusion and speech mode ($p > 0.05$). For the fourth speaker, S1, speech mode was significantly associated with absence of occlusion: 15% (14/90) of the normal tokens and 33% (30/90) of the whispered tokens failed to manifest occlusion. For this speaker, whispered consonants were more likely to be unoccluded [$\chi^2(1,180) = 6.77$, $p < 0.01$].

For speaker S1's whispered tokens, Consonant also had a significant effect on rate of occlusion [$\chi^2(2,90) = 30$, $p < 0.001$]: /d/ accounted for 66.7% (20/30) and /n/ accounted for the remaining 33.3% (10/30) of the unoccluded tokens. All of speaker S1's /t/ tokens manifested occlusion. Voiced alveolars were less likely to manifest occlusion than voiceless alveolars during whisper. Speaker S1 produced at least three occluded repetitions for each VCV item, making it possible to include his data in the repeated measures analysis of variance (Section 3.2).

3.2 Analysis of variance

In order to focus exclusively on the linguopalatal characteristics of occlusion, unoccluded tokens were removed from the following analyses, leaving a total of 625 tokens for examination. Repetitions of each VCV item, normal and whispered, were averaged for each speaker in a repeated measures design. This resulted in 72 data points for each independent variable.

3.2.1 Means

In this section, I will present results relating to the mean values of dependent variables. Speaker was included as a random factor in each linear mixed effects model (LME).

An LME was performed with Duration as the dependent variable and Consonant, Speech Mode, and Vowel as independent variables. The analysis was performed on all /t/ and /d/ tokens. A significant effect was observed for both Consonant (/t/ vs. /d/) and Speech Mode. In both whispered and normal Speech Modes, a greater Duration was observed for /t/ than for /d/ [$F(1,40)=97.29$, $p < 0.001$]. Also, for /t/ and /d/, the Duration was higher in whispered Speech Mode than in normal Speech Mode [$F(1,40)=7.90$, $p < 0.01$]. This indicates that, on average, /t/ is longer than /d/ in any Speech Mode and whispered alveolar stops are longer than normal alveolar stops. Duration results by Speaker for /d/ and /t/ tokens (in all Vowel environments) are shown in Figure 4.

Separate ANOVAs were then performed to determine if any of the individual consonants were significantly longer during whispered speech versus normal speech; we observed that whispered /t/ was longer than normal /t/ [$F(1,19)=28.35$, $p < 0.001$], but that no significant differences were present between whispered and normal /d/ or whispered /n/ and normal /n/.

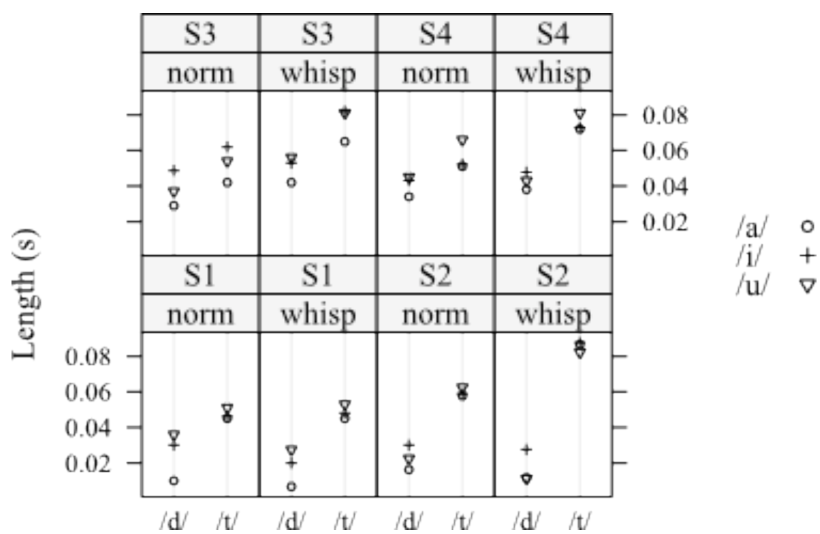

Figure 4: Duration results by Speaker for /d/ and /t/ in each of the three Vowel environments.

The difference could be driven by aerodynamics or perception. Higher air flow could act as a perturbation (in the sense of Lindblom, 1990), causing the speaker to compensate by reducing articulator speed and hyperarticulating. In this case, other evidence of hyperarticulation (such as a lower variability in whispered speech production) might be observed. However, the length difference observed in /t/, but not in /d/, may show that the speaker is making the two articulations more different to enhance cues already present in normal speech: because /t/ is longer than /d/ under normal speech, but longer still under whispered speech, the speaker could be using duration as a cue of phonetic voicing.

An LME was performed with Maximum COG as the dependent variable and Consonant, Speech Mode, and Vowel as independent variables. The analysis was performed on all /t/ and /d/ tokens. A significant effect was observed across all Vowels for Consonant (/t/ versus /d/), but not for Speech Mode. In both whispered and normal Speech Modes, a greater maximum center of gravity was observed for /t/ than for /d/ in all Vowel environments [$F(1,40)=19.95$, $p < 0.001$]. This indicates that, on average, the maximally anterior place of articulation for /t/ is more anterior than it is for /d/. The Maximum COG results for /t/ and /d/ by Speaker are illustrated in Figure 5.

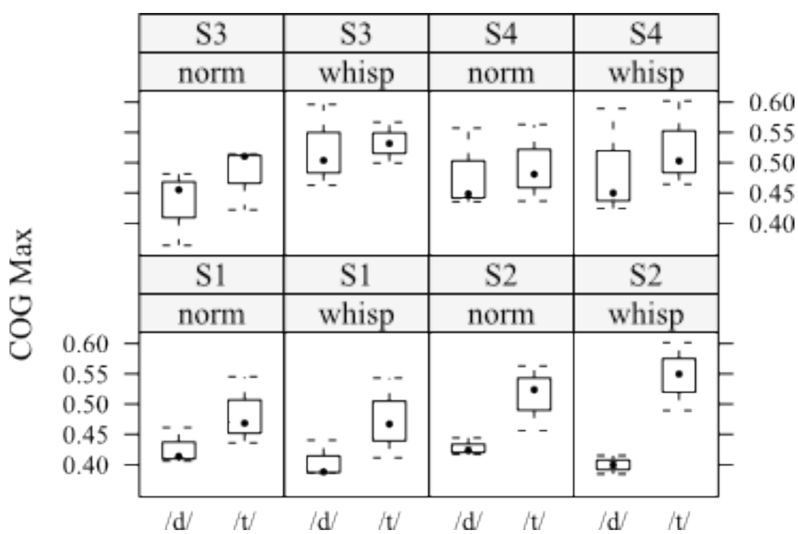

Figure 5: Maximum COG results by Speaker for /d/ and /t/ across all Vowel environments

An LME was performed with Average COG as the dependent variable and Consonant and Speech Mode as independent variables. The analysis was performed on all /t/ and /d/ tokens in the context of the Vowel /u/ (i.e., *udu* vs. *utu*). A significant effect was observed for Speech Mode but not for Consonant. For both /t/ and /d/ in the /u/ context, a greater average center of gravity was observed for whispered tokens than for normal tokens [$F(1,10) = 5.22$, $p < 0.05$]. This indicates that, on average, oral alveolar stops manifest a more anterior place of articulation in whispered speech than they do in normal speech. The average COG results for /d/ and /t/ in the /u/ Vowel environment by Speaker are shown in Figure 6.

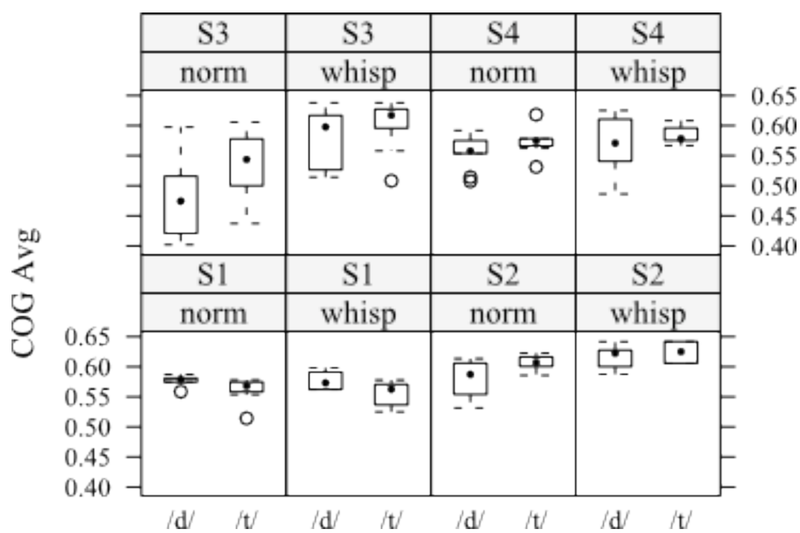

Figure 6: Average COG results by Speaker for /d/ and /t/ in the /u/ Vowel environment.

An LME was performed with Average COG as the dependent variable and Consonant and Speech Mode as independent variables. The analysis was performed on a subset of all /d/ and /n/ tokens in the /u/ Vowel environment. A significant effect was found for Speech Mode, but not Consonant. For both /d/ and /n/, in a /u/ Vowel environment, the average center of gravity was higher in whispered Speech Mode than in normal Speech Mode [$F(1,10)=7.21$, $p < 0.05$]. This indicates that voiced alveolar stops (whether oral or nasal) are articulated farther forward when whispered. The average COG results for /d/ and /n/ in the /u/ Vowel environment by Speaker are given in Figure 7.

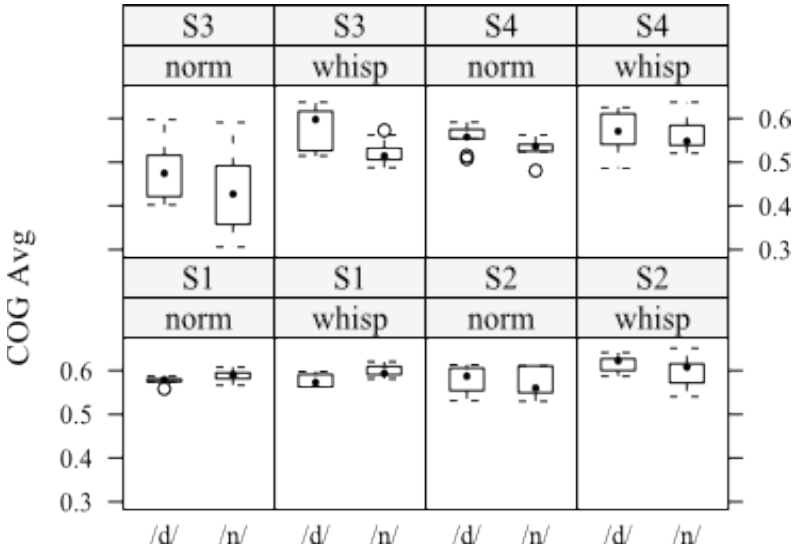

Figure 7: Average COG results by Speaker for /d/ and /n/ in the /u/ Vowel environment.

The Maximum and Average COG results indicate that voiceless stops (/t/) are more anterior than voiced stops (/d/) and that whispered stops are articulated further forward than normal stops. This suggests that aerodynamic considerations play some part in the observed articulatory differences. Both voiceless plosives and whispered plosives are associated with higher rates of air flow (Schwartz, 1972; Stathopoulos et al., 1991; Weismer and Longstreth, 1980) and could be affected by strong egressive speech (cf. Hoole, 1998). The whisper versus normal changes observed seem strongest in the high back vowel (/u/) environment. Possible reasons for this will be discussed further in Chapter 4. It should, however, be mentioned that nothing in the results rules out a perceptual motivation for the observed differences. This, too, will be covered in the discussion.

An LME was performed with SACP as the dependent variable and Consonant and Speech Mode as independent variables. The analysis was performed on a subset of all /t/ and /d/ tokens in the /u/ Vowel environment. A significant effect was found for both Consonant and Speech Mode. For both /t/ and /d/, in a /u/ Vowel environment, a greater average surface area was observed in normal Speech Mode versus whispered Speech Mode [$F(1,10) = 7.84$, $p < 0.05$], indicating that normal alveolar oral stops are articulated with a greater amount of tongue contact than their whispered cognates. Also, across both Speech Modes, /t/ had a greater average surface area than /d/ [$F(1,10) = 38.29$, $p < 0.001$], indicating that unvoiced alveolar plosives are articulated with a greater amount of tongue contact than are voiced plosives. SACP results by Speaker for /d/ and /t/ in the /u/ Vowel environment are given in Figure 8.

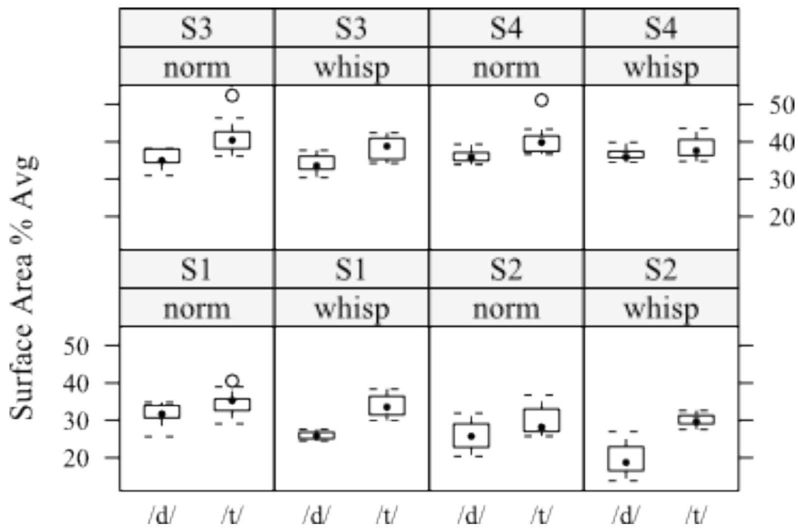

Figure 8: SACP results by Speaker for /d/ and /t/ in the /u/ Vowel environment.

An LME analysis was performed with SACP as the dependent variable and Consonant and Speech Mode as independent variables. The analysis was performed on a subset of all /d/ and /n/ tokens in the /i/ vowel environment. A significant effect was found for both Consonant and Speech Mode. For /d/ and /n/ in an /i/ Vowel environment, a greater average surface area was observed in normal Speech Mode versus whispered Speech Mode, indicating that normal voiced alveolar stops are articulated using a greater amount of tongue contact than their whispered cognates [$F(1,10) = 6.20$, $p < 0.05$]. Also, across both Speech Modes, /n/ had a greater average surface area than /d/, indicating that nasal alveolar stops are articulated with more contact than oral alveolar stops [$F(1,10)=16.96$, $p < 0.01$]. SACP results by Speaker for /d/ and /n/ in the /i/ Vowel environment are shown in Figure 9.

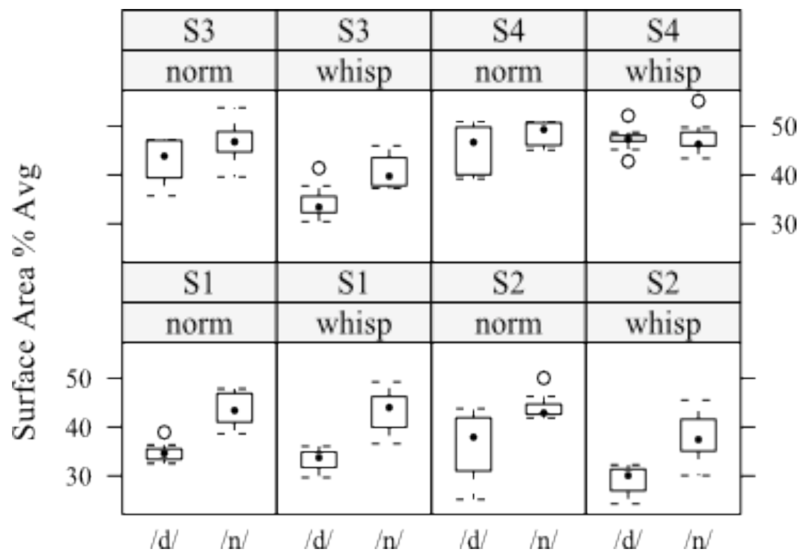

Figure 9: SACP results by Speaker for /d/ and /n/ in the /i/ Vowel environment.

The SACP results indicate that, at least in high vowel environments (/i/ and /u/), normal speech is produced with more tongue-to-palate contact than whispered speech. Additionally, /d/ is articulated with less contact that either /t/ or /n/ in both whispered and normal speech. As the production of /d/ is associated with a greater intraoral pressure than /n/, these results suggest that something other than intraoral pressure may underlie the increased amount of contact exhibited by the /n/. Combined with the variance results which will be discussed next, a greater required precision for whispered speech production (hyperarticulation) may be the cause of the decreased surface area during whisper.

3.2.2 Variance

In this section, I will present statistical results relating to the variability in the data, i.e., the averaged standard deviations between groups of data. The variability in the data should correspond, roughly, to the articulatory (in)stability of the constriction, i.e., how much the constriction changes from frame to frame across its duration.

Average linguopalatal configurations over the duration of the occlusion were compiled for each of the four speakers, as seen in Figures 10-13. In the figures, each cell represents the average, occluded configuration of up to 10 repetitions of the same token (where occlusion occurred). For a single token, e.g. /ana/, black shading indicates that the relevant sensor was contacted during

80% of the occlusion, grey indicates the sensor had contact through 40–80% of the occlusion, and unfilled sensors were contacted less than 40% of the time (conventions following Recasens & Espinosa, 2006). Here, the average configuration of each occlusion has been averaged for up to 10 repetitions of the same token.

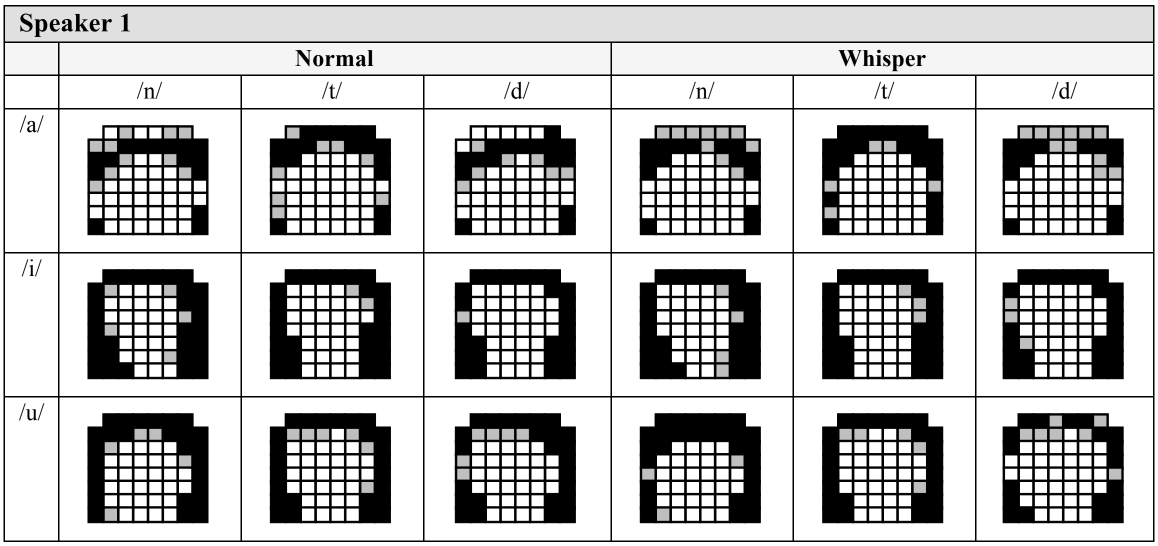

Figure 10: Average linguopalatal configurations during occlusion for Speaker 1. Each cell represents the average, occluded configuration of up to 10 repetitions of the same token (where occlusion occurred). For a single token, e.g. /ana/, black shading would indicate that the relevant sensor was contacted during 80% of the occlusion; grey = 40–80%; unfilled = less than 40% (conventions following Recasens & Espinosa, 2006). Here, the average configuration of each occlusion has been averaged for up to 10 repetitions of the same token.

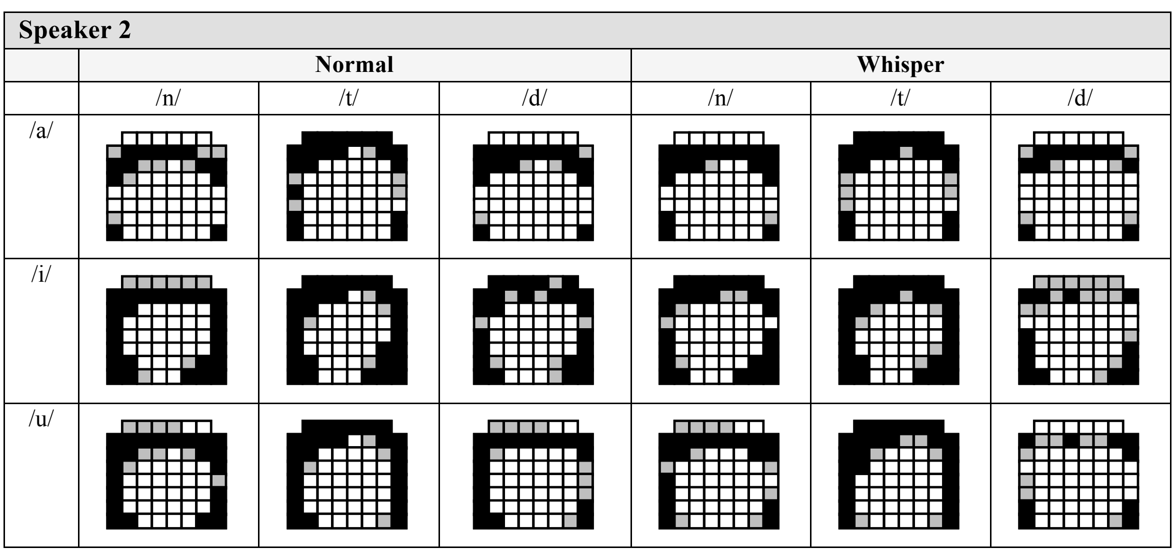

Figure 11: Average linguopalatal configurations during occlusion for Speaker 2. Conventions as detailed in Figure 10.

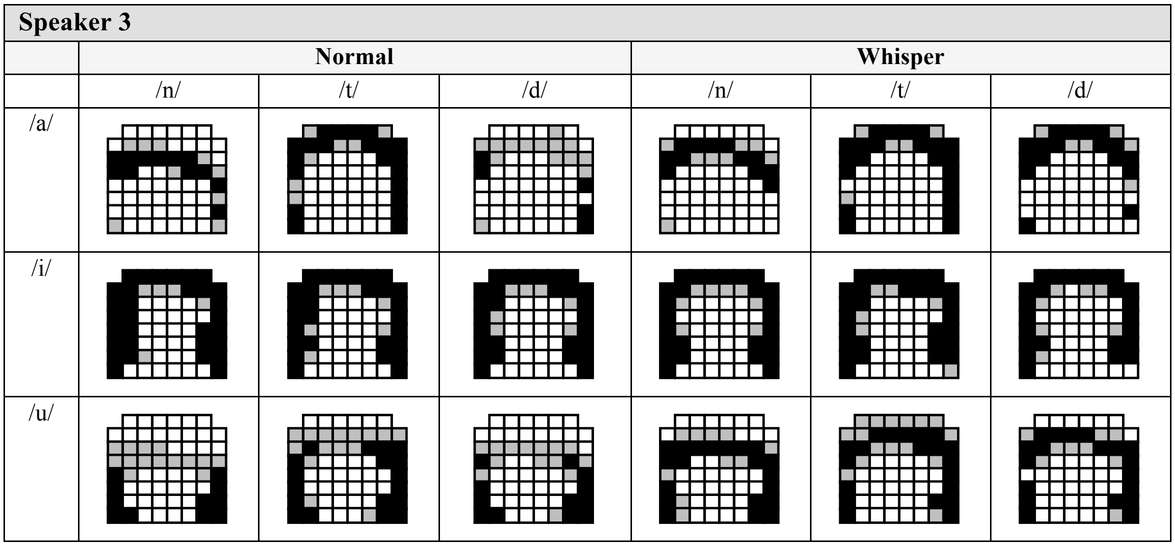

Figure 12: Average linguopalatal configurations during occlusion for Speaker 3. Conventions as detailed in Figure 10.

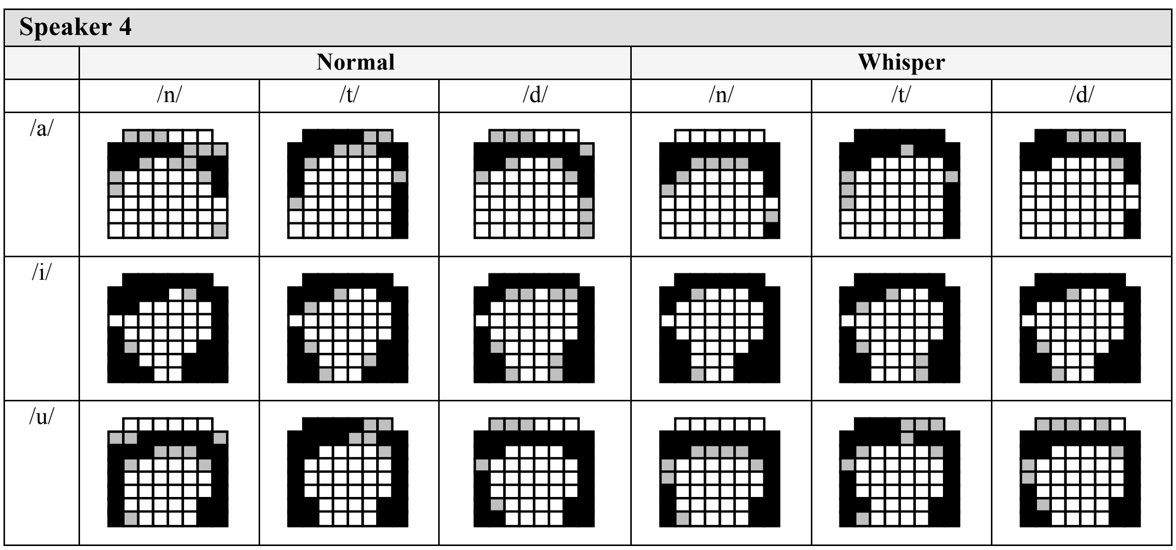

Figure 13: Average linguopalatal configurations during occlusion for Speaker 4. Conventions as detailed in Figure 10.

When examining these figures, differences between normal and whispered speech become apparent. On average, the whispered tokens (on the right side of the figure) show a greater number of frequently-contacted sensors, indicating the higher stability of the speaker's production during whisper.

An LME analysis was performed with standard deviation of the SACP as the dependent variable and Consonant, Mode of speech, and Vowel as independent variables. The analysis was performed on a subset of all /d/ and /t/ tokens. No effects were found by Consonant or Vowel, but an effect was observed for Speech Mode: for /d/ and /t/ across Vowel environments, the *maximum* variability observed was greater in normal Speech Mode than in whispered Speech Mode [$F(1,40)=4.85$, $p < 0.05$]. This indicates that whispered speech is more consistent than normal speech in the amount of linguopalatal contact. The results by Speaker are shown in Figure 14.

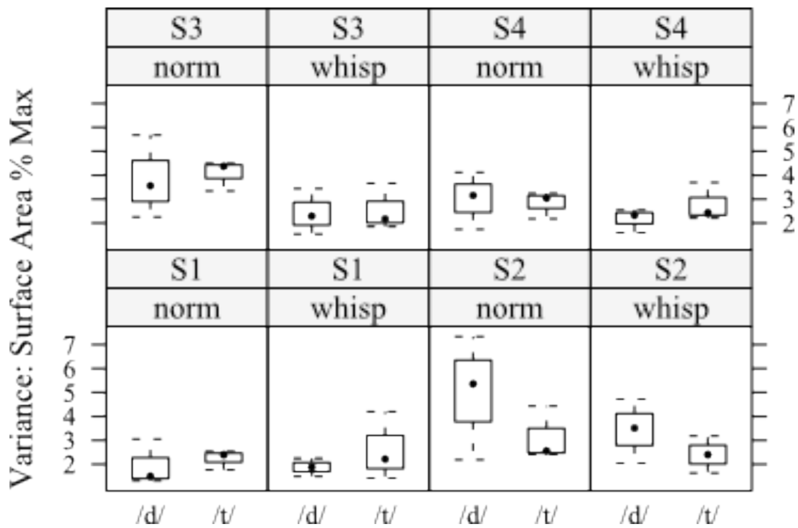

Figure 14: Maximum variance of SACP measures by Speaker for /d/ and /t/ across all Vowel environments. Normal speech shows a higher variability in surface contact (versus whispered speech).

An LME analysis was performed with standard deviation of the SACP as the dependent variable and Consonant, Speech Mode, and Vowel as independent variables. The analysis was performed on a subset of all /d/ and /t/ tokens. No effects were found by Consonant or Vowel, but an effect was observed for Speech Mode: for /d/ and /t/ across the Vowel environments, the *average* variability observed was greater in normal Speech Mode than in whispered Speech Mode [$F(1,40) = 9.70$, $p < 0.01$]. This indicates that whispered speech is more consistent than normal speech in the amount of linguopalatal contact, i.e., whispered speech manifests a greater degree of articulatory stability. The average variance of SACP results are shown by Speaker in Figure 15 below.

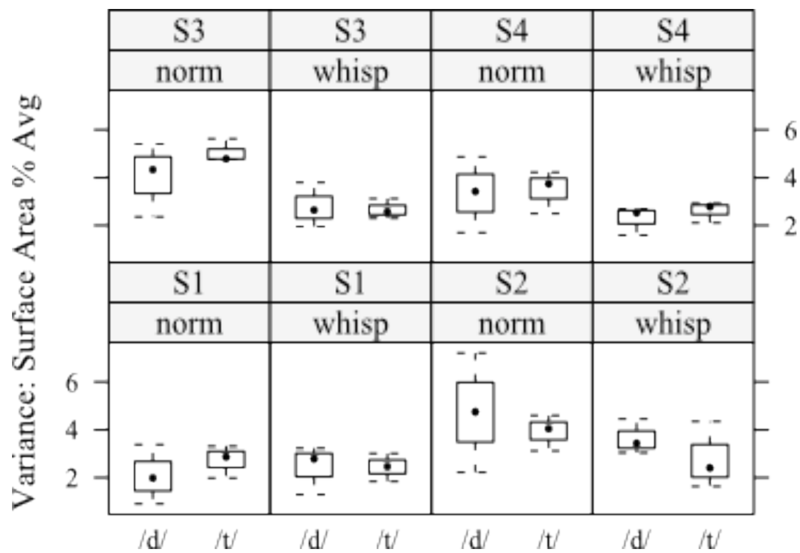

Figure 15: Average variance of SACP measures by Speaker for /d/ and /t/ across all Vowel environments. Normal speech shows a higher variability in surface contact (versus whispered speech).

The variance results demonstrate that, while the voicing of a consonant has no effect on the stability of the occlusion, the mode of speech does: normal speech is consistently performed with more variability (less stability) than whispered speech. This effect extends across all vowel environments and applies to both the maximum variation observed and the average variability across the duration of the occlusion. It appears, then, that whispered speech is associated with hyperarticulation: people may speak more carefully when they whisper to enhance intelligibility.

CHAPTER 4: DISCUSSION

4.1 Glottal and laryngeal effects

Before discussing the findings presented in Chapter 3, I will note the effects of the glottis and larynx on phonemic voicing.

While a perceptual experiment is outside the domain of this study, a study of voicing contrasts necessarily involves discrimination of acoustic cues. While variations in articulation are responsible for acoustic changes that listeners interpret categorically as phonemes, it is not the case that all articulatory changes result in an audible transformation to the speech produced. For example, although some studies have found differences in phonemically voiced and voiceless consonants that persist in whispered speech (Solomon et al., 1989; Mills, 2009), it is unlikely that a glottal or laryngeal gesture along is able to transmit voicing status in whispered speech.

In normal speech, the glottal gesture associated with devoicing of consonants (abduction of the vocal folds), serves to prevent vibration, which inhibits voicing. It has a direct physical effect on the phonetic voicing of the sound. This same glottal gesture (abduction of the vocal folds) is also manifested during whisper. If one speaks of a phonemic devoicing gesture in whisper, it would mean that the vocal folds were abducted by a slightly greater amount. However, additional abduction should have no discernible effect on the "voicing" of the sound, as the vocal folds are already held quite far apart. Further separation should not affect the acoustic output. It is more likely, then, that the laryngeal gestures and glottal differences observed in normal (voiceless) speech and perpetuated during whisper are simply part of the overall motor program associated with phonemic voicing and voicelessness. While an abduction gesture associated with voicelessness in normal speech may persist during whisper, it is unlikely that this gesture effectively transmits a phonemic voicing contrast during whisper.

Because the glottal and laryngeal gestures most likely cannot provide perceptual cues to phonemic voicing during whisper, we must look elsewhere. This is why it is worthwhile to examine the supralaryngeal articulation for differences that may convey this information to the listener.

4.2 Duration

In this study, as in previous work, duration differences were observed both for whispered speech and for voicing contrasts. Examining /t/ and /d/ in normal and whispered speech, we observe two trends: (1) /t/ has a greater duration than /d/ in both speech modes, and (2) whispered oral alveolar stops are, on average, longer than normally-spoken oral alveolar stops. Breaking this down further, we observed that, while the increase in duration during whispered speech observed in /n/ and /d/ was not enough to reach statistical significance, the increase observed for /t/ (17 ms, i.e., +31%) was significant.

These results compare favorably with those of Schwartz (1972), who observed a significant increase in whispered phrases versus normal phrases, and in /p/ but not /m/. However, Schwartz additionally found a significant increase in the duration of /b/, where our results failed to show a change in its analog for the present study, /d/. As in the current study, the largest increase Schwartz observed was that of the voiceless plosive.

The results of the present study also agree to some extent with those of Parnell et al. (1977), who observed a significant duration increase under whisper for /t/, but not for /n/. However, Parnell et al. observed a significant decrease in the duration of /d/ in whispered voice, where the present study revealed a small but non-significant increase. In spite of Parnell et al.'s significant result, they felt the reduction of the /d/ under whisper may have been due to several subjects' apparent "tapping" of /d/, most often during whispered speech. Anecdotally, the tendency toward tapping of /d/, especially in whisper, was also seen in two of the four subjects in the present study, both from the West Coast of the United States. A qualitative examination of the mean productions of each speaker (obviated, in some sense, by the use of a LME model) shows that S1 and S2 tend to produce a whispered /d/ that is shorter than the normal /d/; however, S3 and S4 show at least a slight increase from normal /d/ to whispered /d/. Furthermore, S1 and S2 exhibited whispered /d/ tokens that were as brief as 10ms (= one EPG sample). These differences are reflected in the higher variability that was observed in whispered /d/: the standard deviation is much higher, proportional to its mean, than for whispered /t/ or normal /t/ and /d/.

Very different results were obtained in the present study than those found by Jovičić and Šarić (2008). Limiting comparison to the VCV tokens they collected, we see a mean increase in duration under whispered speech (versus normal speech) of 5.55 ms for /t/, 10.06 ms for /d/, and 9.67 ms for /n/; of these, only the results for /d/ were significant. The discrepancy between the current study's results and those of Jovičić and Šarić is mirrored by Parnell et al. (1977) and Schwartz (1972). Only Schwartz found a significant effect analogous to Jovičić and Šarić's: the increase observed in the voiced oral plosive. One finding that Jovičić and Šarić share with the present study is the relative duration of /t/ to /d/: the authors found that /t/ is longer than /d/ in both normal and whispered speech.

The consistent trends observed in the results as a whole (those of the present study and the others) is that: (1) voiceless stops are longer than voiced stops, in both modes of speech, (2) whispered consonants (at least one consonant for each of the studies examined) tend to be longer than normal consonants, and (3) the voiceless oral stop is the consonant most likely to undergo lengthening in three of the four studies reviewed in Section 1.5.2). Also, there seems to be some evidence (in two of the studies) for a greater proportion of /d/ tokens to be "tapped" in whisper than normal speech, at least for some speakers.

We see, then, some evidence of voiceless and voiced oral stops "moving apart". In other words, the durational gap between /t/ and /d/ is larger in whispered speech than in normal speech. This suggests perceptual intelligibility as a motivating factor in the alternations seen during whisper. While difficulty articulating under the high air flow of whisper may be a cause for some of the lengthening seen at both the consonant and sentential levels, it is unlikely that this can account for the phenomenon entirely. If only aerodynamic differences were at the heart of the durational increase observed during whisper, we might expect this to affect both voiced and voiceless oral stops equally. Instead, we observe the difference between /t/ and /d/ (or /b/ and /p/) increase under whisper. This suggests that speakers try to emphasize the voicing contrast under the degraded acoustic conditions of whisper. As duration has been shown to be a perceptual cue for contrasting voiced and voiceless cognate pairs in normal and whispered speech (Denes, 1955; Dannenbring, 1981; Fuchs et al., 2007), this is not entirely surprising; however, it is interesting that speakers seem able to actively leverage the durational difference they are used to hearing in normal speech for higher intelligibility in their production of whispered speech. Similar findings were reported by Higashikawa et al. (2003), as was detailed in Section 1.5.2, where speakers appeared to leverage an exaggerated volume contrast to differentiate whispered bilabials.

Because /n/ has no voiceless phonemic counterpart in English, and therefore is not predicted to lengthen substantially under whisper, it has been left out of the discussion thus far. However, what would we expect to see in a language that had both voiced and voiceless alveolar nasals? Burmese is one of a few languages that possess a phonemic voiced / voiceless contrasts among nasals (Ladefoged and Maddieson, 1996).[14] This would make it an excellent choice for testing the hypothesis of selective voiceless consonant lengthening. If the durational cue is leveraged by the speaker to help the listener discriminate the whispered voiced from voiceless cognates, we would expect to find the Burmese voiceless nasal significantly lengthen during whisper and the voiced nasal to either resist lengthening or manifest it to a smaller degree. If this is not observed, it would point to another factor (perhaps aerodynamic) causing the observed durational result.

4.3 Center of gravity

Both maximum and average COG during occlusion were measured. Each tells a slightly different story. The tongue, following its loop through the VCV utterance, moves smoothly into and out of the consonant in an elliptical motion (Hoole, 1998). Coarticulation between the consonant and surrounding vowels is substantial: because of the brevity of the utterances and continuous elliptical movement, the tongue is likely to reach its (the tongue's) farthest point for only a few milliseconds. Much of the closure is a midway point between the most anterior articulation of the consonant and the vowels around it.

[14]Jalapa Mazatec (Ladefoged and Maddieson, 1996), Kildin Sami (Kuruch, 1985) Welsh (Jones, 1984), and Central Alaskan Yup'ik (Jacobson, 1995) are languages that maintain (at least) a voiced-voiceless differentiation of the alveolar nasal.

The average COG indicates where the majority of the tongue rests during the entire occlusion. It presents a picture of the coarticulation of the consonant to the surrounding vowels: in other words, it tells us about the VCV utterance it its entirety. Because the contact with the most anterior point is so brief, it figures into the average, but doesn't define it.

The maximum COG, however, is the measure of the most anterior projection of the tongue during the occlusion. While even that point is subject to a coarticulatory effect with the surrounding vowels, it still suggests the limit of the consonant, the configuration of linguopalatal contact when it is most anterior.

We can expect this maximum articulation to mean one of two things, based on the speech condition. In normal speech, a speaker aims at (and usually closely approximates) an articulatory "target" for the consonant he or she is producing (Lindblom, 1990). Because coarticulation is a substantial factor in the tongue's movement during connected speech, it is only likely to strike the target of that consonant (a "true" /d/ or /t/) at the limit of its movement. In this case, during normal speech, the maximum COG represents that articulatory target.

A secondary possibility, especially in perturbed speech, is that the speakers strike somewhere in front of the articulatory target, particularly if air flow is increased substantially (cf. Hoole, 1998). If that were the case, the farthest point reached (the maximum COG) would correspond to this point *past* the target. In either case (the speaker hits the articulatory target or has the tongue move past it), we observe the behavior of the consonant itself as opposed to observing the articulation of the VCV utterance.

Because we are studying the effects of whisper on consonants, and not on VCV utterances, we want to see the behavior at the consonant's target. The limit of the consonant's articulation, the maximum COG, is therefore the more relevant of the two for present purposes.

As was mentioned, /t/ tended to be articulated farther forward than /d/. This occurred in all vowel environments and during both normal and whispered speech. Because the maximum COG was the measure in question, this could indicate the point of the target /d/ and /t/ or, alternatively, a

/d/ and /t/ articulated more anteriorly than intended (in the case of increased air flow). If this (articulation past the target for aerodynamic reasons) occurred, however, it should be observed in the whispered condition (where the air flow is markedly higher) and not during normal speech. Because /t/ had a higher maximum COG in both whispered and normal speech, it is more likely that what was observed is an underlying articulatory difference in the target of /t/ and /d/, one that exists in normal speech and is maintained in whispered speech despite differing aerodynamic conditions. We can reason that the /t/-/d/ distinction is maintained during whispered speech, either because it aids the listener in determining whether the consonant is the phonemically voiced consonant, or because the degree of forward motion is part of each consonant's motor program.

No difference was observed between the average COGs of the consonants. Why is there no difference between consonants in average COG when there was under maximum COG? Perhaps, in normal speech, a small differentiation in tongue placement is not necessary to indicate voicing contrast: the actual introduction of the f0 would overpower any slight difference made by tongue placement. Alternatively, it could be that a correlation exists, but that intra-subject variability is enough to mask this. As will be discussed shortly, the palatal vault shapes of the subjects are those associated with a higher level of variability (Brunner et al., 2009), which could serve to weaken results.

While consonants are not differentiated by measures of average COG, a correlation is seen between modes of speech. For the /u/ environment only, we see alveolar consonants articulated further forward during whispered than normal speech. The aerodynamics may provide the answer. In whisper, the oral air flow is much higher than in normal speech (Schwartz, 1972; Stathopoulos et al., 1991; Weismer and Longstreth, 1980). Hoole found that tongue placement during ingressive stops was more posterior than that of egressive stops. For whisper, which is highly egressive, I report the reverse. That this happens in the /u/ environment make this explanation even more plausible. Because /u/ is a high back vowel, we expect a small constriction before and after the consonant (with respect to /a/); because we are testing alveolars and /u/ is rounded, the constriction would also be smaller before and after the consonant for /u/ than it is for /i/. This tight constriction helps to concentrate the intensified air flow experienced during whisper, thus pushing the tongue forward in its loop (Hoole, 1998). Another explanation for the preference of the /u/ environment for differentiation of movement is that it has been shown to amplify forward movement as compared with the /a/ environment. This has been said to demonstrate that "a redirection of the movement depends on the planning of the target and on the angle of incidence" (Fuchs et al., 2006: 13).

While fricatives and stops have different aerodynamic needs and, thus, do not necessarily act in concert with each other, the differences seen in the results are still worth comparing. Interesting differences can be noted between the present study and prior studies that examine center of gravity and anteriority measures. In opposition to the present study's finding that voiced consonants are posterior to voiceless consonants, McLeod et al. (2003) found that, in normal speech, voiced fricatives tend to be significantly more anterior (as defined by contact over the first two rows) than voiceless fricatives. Just as in the present study, however, no difference in average COG (like that observed by Hardcastle et al., 1991) was found.

Fuchs et al. (2007), like McLeod et al., also showed voiced fricatives to be significantly more anterior than voiceless fricatives in normal speech; however, their findings were evidenced by both higher ANT values (first 4 rows) and higher COG values. These results seem more comparable to the present study's and are more interesting in their contrast. In addition to the different aerodynamic constraints for fricatives and stops, different structural (in regard to tongue bracing) needs and movement patterns exist for the two, as was discussed above in Section 1.5.3 with regard to Fuchs et al.'s (2006) study of stop versus fricative articulation. The differences seen between the present study's results and the results of McLeod et al. (2003) and Fuchs et al. (2007) may further suggest a difference in the movement patterns of stops and fricatives.

One point that bears mentioning is the inclusion of speakers of both sexes within both McLeod et al.'s and Fuchs et al.'s studies. Fuchs et al. found a main effect of speaker sex on productions but included both males and females in their statistical analyses. McLeod et al. also made use of both male and female speakers; while significant inter-speaker variation was found, it was not mentioned whether sex had any significant effect. McLeod et al. did, however, take care to balance the number of male and female speakers of each dialect (each speaker spoke one of five English dialects).

There is evidence that speaker sex can impact the articulation of consonants. Brunner et al. (2009) demonstrated that, while palatal shape (flat versus domed) does not have an effect on acoustic output, it does significantly alter variability in measures of tongue height: speakers with flat palates exhibit less articulatory variability in their productions than those with dome-shaped palates. Flat palatal shapes were further found to be more common in males. Because of the proven gender effects, the present study used only male speakers in order to prevent such confounds. Additionally, when the curvatures of the current study's four speakers' palates were measured using the same procedure and formula outlined in Brunner et al. (as was detailed in Section 2.2), none were found to have a flat palatal shape: two were medium (S1 and S3) and two were domed (S2 and S4). This potentially indicates that a higher degree of variability was seen in these speakers and the effects were less pronounced than they might be with flat-palate speakers.

Another result in contention is the relation of the present study's findings to the AVC (Ohala, 1983). The possibility was posed that this could be seen at the micro level, causing voiced consonants to be articulated farther forward than voiceless consonants, much as voiced consonants are frequently found at anterior points of articulation across languages. The present study does not show this to be the case: while the fronted environment is conducive to voiced consonants, voiced consonants do not appear to require a placement anterior to their unvoiced cognates of the same phonological place.

4.4 Surface area

The SACP results indicate two things: (1) normal speech is produced with more tongue-to-palate contact than whispered speech, at least in high vowel environments, and (2) /d/ is articulated with less contact than either /t/ or /n/ in both whispered and normal speech.

The idea that greater intraoral pressure causes greater surface area contact is based on the premise that the speaker exerts greater lingual pressure on the palate to form a stronger barrier when pressure is higher. The greater lingual pressure would arguably lead to a compression of the tongue, which would result in a greater amount of contact.

Because prior studies are not in agreement on the relationship of intraoral pressure with whispered speech (Schwartz, 1972; Stathopoulos et al., 1991; Murry and Brown, 1976; Weismer and Longstreth, 1980; Kilch, 1982), the whispered versus normal contact differences found in the present study cannot be interpreted as relating to intraoral pressure. However, studies of alveolar stop consonants' intraoral pressure are generally in agreement: /t/ has the greatest intraoral pressure, /n/ has the least, and /d/ falls between the two extremes. Because of this, it should be possible to test our own inter-consonants results against the intraoral pressure findings.

When the current results are compared against the intraoral pressure findings, inconsistencies are seen: specifically, it is problematic that /n/ was found to be articulated with more contact that /d/. Because this is the case, it cannot be that the amount of contact is determined by intraoral pressure.

This does not, however, rule out and effect of lingual pressure; it only means that lingual pressure and intraoral pressure are not (in this case) related. If we assume that the surface area is related to lingual pressure without consideration for intraoral pressure, this indicates that a greater lingual pressure exists in /t/ and /n/ versus /d/. Evidence of this pressure pattern has been found: Brown et al. (1973) observed different lingual pressure orders for each vowel environment. In /iCi/ and /uCu/ sequences, /t/ and /n/ exhibited the highest lingual pressure (/t/ was slightly higher) and /d/ the lowest. However, in /aCa/ sequences, the tongue pressures followed the same pattern as does intraoral air pressure: /t/ to /d/ to /n/.

This parallels the current study's results: in /iCi/ and /uCu/ environments only, /t/ and /n/ both exhibit significantly greater surface area than /d/. However, no significant result was found for consonants in the /aCa/ environment. This was true for both normal and whispered speech, indicating that the lingual pressure ratio between consonants could be part of the set motor programs of the consonants.

This does not, however, explain the why normal speech is articulated with more contact than whispered speech. For this we need to return the discussion of hyper- and hypoarticulation from Section 1.5.4. Following H&H theory (Lindblom, 1990), which suggests that hyperarticulation is associated with target undershoot and decreased articulator velcocities, and Fuchs et al.'s (2006) results, which suggest precise articulation and decreased tongue velocities lead to decreased contact, we would expect hyperarticulation (which is highly precise speech) to exhibit decreased contact.

The present study's findings suggest that normal speech is articulated with more surface area than whispered speech, and that normal speech is faster than whispered speech. This is consistent with the hypothesis that whispered speech is hyperarticulated. Whether to compensate for a degraded speech signal or different aerodynamic conditions, people slow down and speak more carefully when they whisper.

Once again, the results of the present study seem to disagree with those of prior studies (especially studies of fricatives). The present study is unique in reporting greater contact for voiceless than voiced consonants (/t/ versus /d/). McLeod et al. (2003) found no difference in palate contact between /s/ and /z/. Yoshioka (2008) found peak /z/ contact to be greater than peak /s/ contact. Gibbon et al. (2007) found no difference in percent contact between /t/ and /d/. Finally, Fuchs et al. (2007) found that voiced /z/ had more contact (just as it did a higher COG) than voiceless /s/.

There are two main differences between the present study and the others, which may explain the opposing results: a different manner of articulation was studied here than in the others (stops versus fricatives), and a different measure of contact was used. As was just discussed, fricatives seem to have different movement and contact patterns than stops, especially where amount of contact is concerned (Fuchs et al. 2006). Additionally, the aerodynamic requirements of fricatives are vastly different than those of stops. It is not, then, particularly surprising that fricatives and stops would exhibit different results.

The other difference between the prior studies and the present study is the contact measurement used. While the other studies discussed use a measure of contact that is based on the number of sensors contacted, the measure used here is of actual surface area, calculated individually for each subject's palate. As was previously described in the Methods section, the SACP measure eliminates the anterior bias inherent in PC. Comparison between the two measures within the present study showed that, while PC manifested no difference, SACP did. Fuchs et al. (2007) reported to have found only a slight difference in PC between /z/ and /s/, compared with their stronger results for COG and ANT. The authors' PC measure happens to pattern with their ANT measure. It is thus highly possible that the increased anteriority exhibited by /z/ presented enough of a bias in the PC measure to show a significant difference. Furthermore, two of her subjects actually manifested the reverse pattern (/s/ > /z/), even with the extra anterior weight given by the PC.

The methodological differences between the present study and each of the others, as well as the different articulatory patterns attested for stop and fricative consonants, make the inconsistencies themselves in the results less suggestive of an interesting pattern.

4.5 Variance

The present study did find two results which fit well with and inform a number of the other studies discussed. One theory raised by many of these authors is that of the "difficulty" of whispered speech. Parnell et al. (1977) felt speakers slowed down to increase intelligibility. Jovčić and Šarić (2003) also suggested that whisper was more difficult because of the increased air flow. Furthermore, they asserted that whichever consonants experienced the greatest time increases during whisper must be the hardest consonants to articulate: their results showed that palatals were especially problematic and take a greater degree of articulatory accuracy to achieve. While the authors suggested that evidence of this idea was present in their duration findings, they felt that further research was needed to determine the veracity of their hypothesis. However, while there is ample evidence that speech is slower during whisper (Jovčić and Šarić, 2003; Parnell et al., 1977; Schwartz, 1972; Schwartz, 1968), is there any indication that whisper is more carefully produced (i.e., hyperarticulated)?

Variance, according to Lindblom's (1990) H&H theory discussed above, is a way to measure where an instance of speech fits along the spectrum of hyperarticulation and hypoarticulation. In the H&H theory, when a speaker shifts to a type of speech where output constraints dominate (like loud speech), the speaker's focus on that target results in hyperarticulation, and a decrease in articulatory variance is seen.

The variance results in the present study demonstrate that, while the consonant spoken has no effect on the stability of the occlusion, the mode of speech does: normal speech is consistently performed with more variability (less stability) than whispered speech. This effect extends across all vowel environments and applies to both the maximum variation observed, as well as the average variability across the duration of the occlusion. Figures 10-13 provide visual evidence of this. On average, the whispered tokens (on the right side of the figure) show a greater number of frequently-contacted sensors, indicating the higher stability of the speaker's production during whisper.

The decrease in variance observed during whispered speech supports the idea that whisper acts as perturbation (as in H&H theory). As a perturbation, whisper asserts output constraints (for example, different aerodynamic conditions and reduced intelligibility to the listener) on speech, requiring compensatory actions by the speaker and causing hyperarticulation. In other words, as suggested by Jovčić and Šarić (2003), people speak more carefully when whispering.

CHAPTER 5: CONCLUSION

Supralaryngeal articulatory differences observed in whispered versus normal speech are likely motivated by the perceptual and aerodynamic requirements with which whisper is associated. There is evidence that speakers actively increase the duration of unvoiced consonants (like /t/) during whisper to accentuate a voiced-voiceless length difference already present in normal speech. An increase in maximum center of gravity in whispered (versus normally-spoken) consonants and in phonemically unvoiced (versus phonemically voiced) consonants seems to be a passive articulatory difference motivated by higher air flow present during whisper. Compensatory hyperarticulation of whispered speech is indicated by surface area differences between normal and whispered speech: whispered speech shows less contact than normal speech, a situation indicative of lower articulator velocity and more careful placement of the tongue on the palate (higher precision movement). In one speaker, absence of occlusion was associated with whisper. Finally, further evidence of hyperarticulation during whisper is present in the variance findings: lower variability (more stability) is present in whispered speech than in normal speech, a difference that was observed across all vowels. While passive aerodynamic changes are observed, the evidence of hyperarticulation confirms, to a certain extent, the Intelligibility hypothesis: consciously or not, speakers hyperarticulate during whisper to increase or maintain the comprehensibility of their speech.

REFERENCES

Articulate Instruments Ltd. (2008a). *Articulate Assistant User Guide V.1.17*. Musselburgh, UK: Articulate Instruments Ltd.

Articulate Instruments Ltd. (2008b). *WinEPG Installation and User's Manual* (revision 1.16 ed.). Musselburgh, UK: Articulate Instruments Ltd.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Boston, MA: College-Hill Press.

Baken, R., & Orlikoff, R. (2000). *Clinical Measurement of Speech and Voice*. San Diego, CA: Singular.

Barnes, L., Tse, L. L. Y., Hunt, J. L., Brandwein-Gensler, M., Urken, M., Slootweg, P., Gale, N., Cardesa, A., Zidar, N., & Boffetta, P. (2005). Tumours of the hypopharynx, larynx, and trachea: Introduction. In L. Barnes, J. W. Eveson, P. Reichart, & D. Sidransky. (Eds.) *World Health Organization Classification of Tumors. Pathology and Genetics of Head and Neck Tumours* (pp. 111-121). Lyon: IARC Press.

Brown, W. S., McGlone, R. E., & Proffit, W. R. (1973). Relationship of Lingual and Intraoral Air Pressures during Syllable Production. *Journal of Speech and Hearing Research*, *16*, 141-151.

Brunner, J., Fuchs, S., & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. *Journal of the Acoustical Society of America*, *125*(6), 3936–3949.

Butcher, A. (1989). Measuring coarticulation and variability in tongue contact patterns. *Clinical Linguistics and Phonetics*, *3*(1), 39-47.

Chang, C. B. & Yao, Y. (2007). Tone production in whispered Mandarin. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1085-1088, Saarbrücken: Universität des Saarlandes.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *3*, 466–485.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Dannenbring, G.L. (1980). Perceptual discrimination of whispered phoneme pairs. *Perceptual and Motor Skills*, *51*(3), 979–985.

Denes, P. (1955). Effect of duration on the perception of voicing. J*ournal of the Acoustical Society of America*, *27*, 761-764.

Esling, J. H. (1984). Laryngographic study of phonation type and laryngeal configuration. *Journal of the International Phonetic Association*, *14*, 56-73.

Esling, J. H. & Harris, J. G. 2003. An expanded taxonomy of states of the glottis*. Proceedings of the 15th International Congress of Phonetic Sciences*, 1049-1052. Barcelona, Spain: UAB.

Fontdevila, J., Pallarè, M. D., & Recasens, D. (1994). The contact index method of electropalatographic data reduction, *Journal of Phonetics*, *22*, 141-154.

Fuchs, S., Perrier, P., Geng, C., & Mooshammer, C. (2006). What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. In J. Harrington & M. Tabain (Eds.), *Towards a better understanding of speech production processes* (pp. 149-164). New York: Psychology Press.

Fuchs, S., Brunner, J., & Busler, A. (2007). Temporal and spatial aspects concerning the realizations of the voicing contrast in german alveolar and postalveolar fricatives. *Advances in Speech Language Pathology*, *9*(1), 90-100. Author Transcript version retrieved from hal.archives-ouvertes.fr/docs/00/37/13/21/PDF/Fuchs_et_al_-_to_T_F.pdf

Fujimara, O., & Miller, J. E. (1979). Mandible height and syllable-final tenseness. *Phonetica*, 36(4-5), 263-272.

Gibbon, F., Yuen, I., Lee, A., & Adams, L. (2007). Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *Advances in Speech and Language Pathology*, *9*, 82–89.

Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research*, 37(1), 4-27.

Greene, M. C. L. (1980). *The Voice and its Disorders* (4th ed.). Philadelphia: J. P. Lippencott.

Granit, R. (1979). *The Purposive Brain*. Cambridge, MA: MIT Press.

Hardcastle, W. J. (1972). The use of electropalatography in phonetic research. *Phonetica*, *25*, 197–215.

Hardcastle, W. J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, *19*, 251–266.

Higashikawa, M., Green, J., Moore, C. A., & Minifie, F. D. (2003). Lip kinematics for /p/ and /b/ production during whispered and voiced speech. *Folia Phoniatrica et Logopaedica*, *55*, 17-27.

Houde, R.A. (1968). A study of tongue body motion during selected speech sounds. *Speech Communication Research Laboratory, Santa Barbara, Monograph No. 2*. Santa Barbara, CA.

Hoole, P. (1998). Do Airstream Mechanisms Influence Tongue Movement Paths? *Phonetica*, *55*(3), 131-146.

Ingham, R. J., Bothe, A. K., Jang, E., Yates, L, Cotton, J., and Seybold, I. (2003). Measurement of Speech Effort During Fluency-Inducing Conditions in Adults Who Do and Do Not Stutter. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1286-1301.

Jacobson, S. (1995), *A Practical Grammar of the Central Alaskan Yup'ik Eskimo Language*, Fairbanks: Alaska Native Language Center.

Jones, G. E. (1984). The distinctive vowels and consonants of Welsh. In M. J. Ball & G. E. Jones (Eds.), *Welsh Phonology: Selected Readings* (pp. 30-64). Cardiff: University of Wales Press.

Jovičić, S. T., & Šarić, Z. (2008). Acoustic analysis of consonants in whispered speech. *Journal of Voice*, 22, 263-274.

Kent, R., & Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. *Journal of Speech, Language, and Hearing Research*, *15*, 453–473.

Klich, R. J. (1982). Effects of speech level and vowel context on intraoral air pressures in vocal and whispered speech. *Folia Phoniatrica*, *34*, 33-40.

Koenig, L. L., Lucero, J. C., & Mencl, W. E. (2008). Laryngeal and aerodynamic adjustments for voicing versus devoicing of /h/: A within-speaker study. *Journal of Voice, 22*(6), 709-720.

Kuruch, R. (1985). (in Russian*) A brief grammatical sketch of the Sami language*. Moscow: Mokwa.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lim, B. P. (2010). *Computational Differences in Whispered and Non-Whispered Speech*. Doctoral dissertation, University of Illinois at Urbana-Champaign, unpublished draft of November 29, 2010.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). The Netherlands: Kluwer Academic Publishers.

Lindqvist-Gauffin, J. (1969). Laryngeal mechanisms in speech. *Department for Speech, Music and Hearing Quarterly Progress and Status Report*, *10*(2-3), 026-032.

Liu, S., & Samuel, A.G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*, 109-138.

Löfqvist, A., Baer, T., McGarr, N.S., & Seider-Story, R. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America, 85*, 1314-1321.

Löfqvist, A., & Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research*, 40(4), 877-893.

Löfqvist, A., and McGarr, N. S. (1987). Laryngeal dynamics in voiceless consonant production. In T. Baer, C. Saski, & K. S. Harris (Eds.), *Laryngeal Function in Pronation and Respiration* (pp. 391-402). Boston: College-Hill.

McAuliffe, M. J. & Ward, E. C. (2006) The use of electropalatography in the assessment and treatment of acquired motor speech disorders in adults: Current knowledge and future directions. *NeuroRehabilitation, 21*(3), 189-203.

McLeod, S. (2006). Australian adults' production of /n/: An EPG investigation. *Clinical Linguistics and Phonetics*, *20*, 99–107.

McLeod, S., Robers, A., & Sita, J. (2003). The difference between /s/ and /z/: More than +/- voice? Poster presented at American Speech-Language-Hearing Association Convention, Chicago, IL, November.

Mills, T. I. P. (2003). *Cues to voicing contrasts in whispered Scottish obstruents*. M. A. Thesis, University of Edinburgh. Retrieved from http://www.lel.ed.ac.uk/~tmills/files/TimothyMillsMSc.pdf

Mills, T. I. P. (2009). *Speech motor control variables in the production of voicing contrasts and emphatic accent*. Doctoral dissertation, University of Edinburgh. Retrieved from www.ling.ed.ac.uk/~tmills/files/TimothyMillsPhD.pdf

Monoson, P., & Zemlin, W. R. (1984). Quantitative study of whisper. *Folia Phoniatrica*, *36*, 53-65.

Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P., & Fuchs, S. (2003). Coordination of lingual and mandibular gestures for different manners of articulation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 81-84. Barcelona, Spain: UAB.

Murry, T., & Brown, W. (1976). Peak intraoral air pressures in whispered stop consonants. *Journal of Phonetics*, *4*, 183-187.

Nicholson, H., & Teig, A.H. (2003). How to tell beans from farmers: cues to the perception of pitch accent in whispered Norwegian. *Nordlyd*, *31*(2), 315-325.

Nguyen, N. (1997). loadepg.m. Available at http://www.articulateinstruments.com/EMATools2.zip. Downloaded July 31, 2009.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189-216). New York: Springer-Verlag.

Parnell, M., Amerman, J. D., & Wells, G. B. (1977). Closure and constriction duration for alveolar consonants during voiced and whispered speech. *Journal of the Acoustical Society of America*, *86*, 1678-1683.

Perkell, J.S. (1969). *Physiology of speech production: results and implications of a quantitative cineradiographic study*. Cambridge: MIT Press.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434-446.

Recasens, D., & Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan, *Journal of Phonetics*, *34*, 295-318.

Recasens, D., Fontdevila, J. & Pallarés, D. (1995), Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. *Journal of Phonetics*, *24*, 165-185.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*, 173-189.

Schwartz, M. F. (1968). Effect of masking noise upon syllable duration in oral and whispered reading. *Journal of the Acoustic Society of America*, *43*, 169–170.

Schwartz, M. F. (1972). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *Journal of the Acoustical Society of America*, *51*, 2025-2029.

Sherrington, C. S. (1941). *Man on His Nature*. London: MacMillan.

Shosted, R. K. (accepted). An articulatory-aerodynamic approach to stop excrescence. *Journal of Phonetics*.

Shosted, R. K. (in prep.). An anatomically-based surface area measure of linguopalatal contact.

Solomon, N. P., McCall, G. N., Trosset, M. W., & Gray, W. C. (1989). Laryngeal configuration and constriction during two types of whispering. *Journal of Speech, Language, and Hearing Research*, *32*, 161-174.

Stathopoulos, E. T., Hoit, J. D., Hixon, T. J., Watson, P. J., & Solomon, N. (1991). Respiratory and Laryngeal Function during Whispering. *Journal of Speech and Hearing Research*, *34*(4), 761-767.

Stone M., Faber A., Raphael L., Shawker T. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics*, *20*, 253–270.

Sussman, H. M., MacNeilage P.F., and Hanson R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: preliminary observations. *Journal of Speech, Language, and Hearing Research*, *16*(3), 397-420.

Tanokuchi, F., Sakai, S., Kawano, M., & Isshik, N. (1986). Articulation Training for Velopharyngeal Function Reinforcement. *Studia Phonologica*, *20*, 38-48.

Tartter, V. C. (1989). What's in a whisper? *Journal of the Acoustical Society of America*, *86*, 1678-1683.

Tartter, V. C. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America*, *96*(4), 2101-2107.

Weismer, G., & Longstreth, D. (1980). Segmental gestures at the laryngeal level in whispered speech: Evidence from an aerodynamic study. *Journal of Speech, Language, and Hearing Research*, 23(2), 383-392.

Yoshioka, H. (2008). The role of tongue articulation for /s/ and /z/ production in whispered speech. *Acoustics '08 Paris Proceedings*, 2335-2338. Paris: Societé Française d'Acoustique.

Zemlin, W. R. (1998). *Speech and Hearing Science: Anatomy and Physiology*. (4th ed.). Boston, MA: Allyn & Bacon.

Zeroual, C., Esling, J. H., & Crevier-Buchman, L. (2005). Physiological study of whispered speech in Moroccan Arabic. *Interspeech 2005 Proceedings*, 1069-1072. Lisbon.

APPENDIX

List of all **token** items, in the *carrier phrase*, with distractors denoted by *:

Say **aba** *again**

Say **ada** *again*.

Say **ama** *again**

Say **ana** *again*

Say **apa** *again**

Say ata *again*

Say **ibi** *again**

Say **idi** *again*

Say **imi** *again**

Say **ini** *again*

Say **ipi** *again**

Say **iti** *again*

Say **ubu** *again**

Say **udu** *again*

Say **umu** *again**

Say **unu** *again*

Say **upu** *again**

Say **utu** *again*